THE UNIVERSITY *of* EDINBURGH

Biophysical Studies into the Structure and Interactions of Proteins and Peptides

Sophie Rebecca Harvey

Ph.D.

The University of Edinburgh

2014

'There is nothing like looking, if you want to find something…You certainly usually find something, if you look, but it is not always quite the something you were after.'

-J.R.R.Tolkien
The Hobbit

Declaration

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Edinburgh. This work has not been submitted for any other degree or professional qualification. The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 1: This introductory chapter includes information discussed in a review article; *Ion mobility mass spectrometry for peptide analysis,* ***Sophie R. Harvey****, Cait E. MacPhee and Perdita E. Barran, Methods,* ***2011****, 54, 454-461*. My contribution to this paper is as follows; literature searching for and reading of relevant articles as well as authoring the manuscript and preparing figures.

Chapter 3: This chapter is based upon work from a jointly-authored publication; *Small-molecule inhibition of c-MYC:MAX leucine zipper formation is revealed by ion mobility mass spectrometry,* ***Sophie R. Harvey****, Massimiliano Porrini, Christiane Stachl, Derek MacMillan, Giovanna Zinzalla and Perdita Barran, JACS,* ***2012****, 134, 19384-19392*. My contribution to this work is as follows; the design and implementation of all MS and DT IM-MS experiments, collection, analysis and interpretation of data, authoring the manuscript, and preparing all figures. Molecular dynamics (MD) simulations were performed by Dr Massimiliano Porrini.

Chapter 4: This chapter includes results which form the basis of a recently submitted paper; *Dissecting the Dynamic Conformations of the Metamorphic Protein Lymphotactin.* ***Sophie R. Harvey****, Massimiliano Porrini, Albert Konijnenberg, David J. Clarke, Robert C. Tyler, Cait E. MacPhee, Patrick R. R. Langridge-Smith, Brian F. Volkman and Perdita E. Barran, in review.* My contribution to this work is as follows; design and implementation of all MS and DT IM-MS experiments, collection of data and the subsequent analysis and interpretation. Furthermore, I collected and analysed all ECD data, in addition to authoring the manuscript, and preparing figures. MD simulations were performed by Dr Massimiliano Porrini.

Chapter 6: MD simulations discussed in this chapter were performed by Dr Massimiliano Porrini.

……

Sophie Rebecca Harvey
April 2014

Acknowledgments

I would firstly like to thank my amazing supervisors Professor Perdita Barran and Professor Cait MacPhee without whom the work presented here would not have been possible. Thank you both so much for your continued help, support, encouragement and advice.

I have had the pleasure of working with great collaborators. Thank you to Derek Macmillan for providing the c-MYC and MAX peptide samples, and to Giovanna Zinzalla for her insight into this system. Thanks also go to Brian Volkman, Rob Tyler and Jamie Fox for providing lymphotactin samples and for the interesting discussions. I am grateful to Dusan Uhrín for providing glycosaminoglycan samples and to Julia Dorin for insights into defensins.

I am lucky to have shared my time in the lab with a great group of people. Thank you for the help, chats, group outings, numerous cups of coffee and cake; Martin, Roland, Jude, Yana, Hattie, Ewa, Kamila, Max, Jason, Albert, Jakub, Bex, Chris, Ashley, Ellie and Alex. I am also grateful to Albert and Chrissy for their help with lymphotactin and c-MYC. A big thank you also to Max, for all his help with molecular modelling.

I would like to acknowledge Dave Clarke and Logan MacKay for their advice and help with the ECD experiments. I am also grateful for all the technical help and support from Donald, Stuart and Davey in the chemistry workshops and Hunter and Richard in the physics electronic workshop. Thanks also to the great teams in the chemistry stores and admin office. Thank you to Janice Bramham and Liz Blackburn for ITC and CD training, and to Steve Mitchell for his help with TEM.

A thank you also to my friends for the many laughs, cups of tea, girls nights, brunches, dances and distractions when needed. A special mention to Alison for her understanding, advice, snacks and notes, thank you for all the great chats shared over noodles and ginger beer.

Finally, to my parents I am so grateful for your constant love and support. To my sister Vicki, thank you for always being there and for listening. To my brother Adam, thanks for always finding the time to encourage, advise and provide a different perspective on things. Also thanks to the four legged members of the Harvey family, for letting me drag you out for writing break walks when you would have much rather been snoring by the fire.

Abstract

Investigating the structure of proteins and their interactions with other biomolecules or drug molecules, coupled with the consideration of conformational change upon binding, is essential to better understand their functions. Mass spectrometry (MS) is emerging as a powerful tool to study protein and peptide structure and interactions due to the high dynamic range, low sample consumption and high sensitivity of this technique, providing insight into the stoichiometry, intensity and stability of interactions. The hybrid technique of ion mobility-mass spectrometry (IM-MS) can provide insight into the conformations adopted by protein and peptide monomers and multimers, in addition to complexes resulting from interactions, which when coupled with molecular modelling can suggest candidate conformations for these *in vacuo* species and by inference their conformations in solution prior to ionisation and desolvation. The work presented in this thesis considers a number of different peptide and protein systems, highlighting how the combination of MS and IM-MS based techniques, in conjunction with other biophysical techniques such as circular dichroism (CD) spectroscopy, transmission electron microscopy (TEM) and isothermal titration calorimetry (ITC) can provide insight into these dynamic systems.

First a case study into the ability of MS and IM-MS to study disorder-to-order transitions is presented. The transcription factor c-MYC can only perform its function upon binding with its binding partner MAX; deregulation of c-MYC is, however, implicated in a number of human cancers. c-MYC and MAX comprise intrinsically disordered regions which form a leucine zipper upon binding. The work presented here focuses on the leucine zipper regions of both c-MYC and MAX, their individual conformations and changes upon binding. Inhibiting the c-MYC:MAX interaction is a current target for drug therapy and hence the inhibition of this interaction with a previously identified small drug-like molecule was also examined using these techniques, to determine if such an approach may be appropriate for investigation of future therapeutics.

Next the ability of MS-based techniques to preserve, transmit and distinguish between multiple conformations of a metamorphic protein was examined. The chemokine lymphotactin has been shown to exist in two distinct conformations in equilibrium in a ligand-free state. The existence of such metamorphic proteins has called into question whether traditional structural elucidation tools have been inadvertently biased towards consideration of single conformations. Here, the potential of gas-phase techniques in the study of conformationally dynamic systems is examined through the study of wild type

lymphotactin and a number of constructs designed either as a minimum model of fold or to mimic one of the distinct folds.

Interactions between chemokines and glycosaminoglycans (GAGs) are thought to be essential for the *in vivo* activity of these proteins. The interactions between the distinctive chemokine lymphotactin and a model GAG were hence probed. As with the structural studies, additional protein constructs were considered either to represent the minimum model of fold, one distinct fold of the metamorphic protein or designed to diminish its GAG binding propensity. The ability of each construct to bind GAGs, the stoichiometry of the interactions and conformations adopted by the resulting complexes in addition to aggregation occurring upon the introduction of the GAG is considered.

Finally, the similarities, with respect to structure and function, between the chemokine superfamily of proteins and the human β-defensin subfamily of antimicrobial peptides are considered. The tendency of human β-defensins 2 and 3 to bind a model GAG is examined; the stoichiometry of binding and conformations adopted and aggregation occurring here are considered and compared with that of chemokines.

Lay Summary

Proteins and peptides are essential components of all living organisms which carry out specific roles and are crucial for the correct functioning of organisms. They are composed of chains of building blocks known as amino acids; there exist 20 standard natural amino acids, however, the sequence of amino acids varies depending on the protein or peptide. Peptides and proteins are distinguished based on their size with peptides being smaller, composed of fewer than 50 amino acids. The sequence of amino acids enables species to adopt specific three-dimensional folds, which can contain regions of order and disorder, and enables peptides and proteins to carry out their functions. Proteins are not rigid moieties and instead can undergo conformational changes depending on their environment or in response to binding with specific partners, often in order to carry out their function. In some cases, the so-called metamorphic proteins, proteins can even exist in multiple conformations in equilibrium. The study of protein and peptide conformations and conformational changes is, therefore, essential in order to better understand their functions.

In order to carry out their functions proteins and peptides often have to bind other species such as proteins, peptides, biomolecules or ligands. The difficulty in studying such systems often comes from being able to differentiate between the different components in a heterogeneous mixture. For example, to study two peptides that can interact, one has to be able to clearly distinguish between the two free peptides and the complexes occurring. The work presented in this thesis uses a combination of techniques to study such systems which can distinguish different components based on their mass (using a technique known as mass spectrometry) and on their mass and conformation (using a technique called ion mobility-mass spectrometry), and additional characterisation techniques where appropriate. The added advantage of such techniques is they require very little sample. The applicability of these techniques to study a variety of questions is probed. Shape-changing proteins and peptides are first considered to determine if these techniques could distinguish between the different conformations and complexes. Two peptides which exist in unstructured forms until they bind each other, forming a helical complex, were first studied. Next a protein known to exist in two distinct conformations in equilibrium, in a ligand-free state was considered.

A number of proteins and peptides are known to interact with sugars (glycosaminoglycans) which are either present on the surface of cells or in a region between cells, the extracellular matrix. Studying the complexes between proteins/peptides and sugars can be difficult, however, as in the presence of glycosaminoglycans proteins/peptides can bind the sugar, interact with themselves forming multimers or even glycosaminoglycan bound multimers,

and hence can result in heterogeneous mixtures. Here, the ability of mass spectrometry to distinguish between the different species present, and identify complex formation, is studied for both proteins and peptides. Furthermore, the conformations adopted upon binding are determined using ion mobility-mass spectrometry, shedding light on this complicated, biologically important process.

Table of contents

Appendix

Abbreviations

Ala	alanine
AmAc	ammonium acetate
Arg	arginine
Asp	aspartic acid
ATD	arrival time distribution
b-HLH-Zip	basic helix-loop-helix leucine zipper
BPTI	bovine pancreatic trypsin inhibitor
CC3	Ltn monomer mutant with additional disulfide bond
CCL	CC-type chemokine (ligand)
CCR	CC-type chemokine receptor
CCS	collisional cross section
CD	circular dichroism
CID	collision induced dissociation
CSD	charge state distribution
CTD	C-terminal domain
CXCL	CXC-type chemokine (ligand)
CXCR	CXC-type chemokine receptor
Cys	cysteine
Da	Dalton
DC	direct current
DE_X	decameric protein species, where x is the abbreviation of the protein
DNA	deoxyribonucleic acid
DT	drug target region
DT IM-MS	drift tube ion mobility-mass spectrometry
D_X	dimeric protein species, where x is the abbreviation of the protein
ECD	electron capture dissociation
EHSS	exact hard sphere scattering
ESI	electrospray ionisation
FAIMS	field asymmetric ion mobility spectrometry
FT ICR-MS	Fourier transform ion cyclotron resonance mass spectrometer
Fx	fondaparinux
GAG	glycosaminoglycan
Glu	glutamic acid

GPCR	G protein-coupled receptor
HBD	human β-defensin
HDMS	high definition mass spectrometer
HLH	helix-loop-helix
H_X	hexameric protein species, where x is the abbreviation of the protein
ICR	ion cyclotron resonance
ID	intrinsically disordered
IDP	intrinsically disordered protein
IM-MS	ion mobility-mass spectrometry
IMS	ion mobility spectrometry
IR	infrared spectroscopy
ITC	isothermal titration calorimetry
Ltn	lymphotactin
Ltn10	monomeric fold of Ltn, solved at 10 °C
Ltn40	dimeric fold of Ltn, solved at 40 °C
Lys	lysine
m/z	mass-to-charge ratio
MALDI	matrix assisted laser desorption ionisation
MCP	microchannel plate
MD	molecular dynamics
m-NBA	*meta*-nitrobenzyl alcohol
MOBCAL	a program to CALculate MOBilities
MRE	mean residue ellipticity
mRNA	messenger ribonuleic acid
MS	mass spectrometry
M_X	monomeric protein species, where x is the abbreviation of the protein
NaCl	sodium chloride
NaI	sodium iodide
n-ESI	nano-electrospray ionisation
NLS	nuclear localisation signal
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
O_X	octameric protein species, where x is the abbreviation of the protein
PA	projection approximation
PDB	protein data bank

PSA	projection superposition approximation
P_X	pentameric protein species, where x is the abbreviation of the protein
Q-TOF	quadrupole time-of-flight mass spectrometer
R23A	mutant of Ltn in which the arginine residue 23 is mutated to alanine
R23A/R43A	mutant of Ltn in which the arginine residues 23 and 43 are mutated to alanine
R43A	mutant of Ltn in which the arginine residue 43 is mutated to alanine
RF	radio frequency
RMSD	root-mean-square deviation
RNA	ribonuleic acid
SA	simulated annealing
TAD	transactivation domain
TDC	time-to-digitial convertor
TEM	transmission electron microscopy
TE_X	tetrameric protein species, where x is the abbreviation of the protein
TM	trajectory method
TOF	time-of-flight
TWIG IM-MS	travelling wave ion guide ion mobility-mass spectrometry
T_X	trimeric protein species, where x is the abbreviation of the protein
UA	uranyl acetate
W55D	Ltn dimer mutant, trypyophan residue 55 is mutated to aspartic acid
WT	wild type
WT 1-72	construct of wild type Ltn consisting of residues 1-72
WT 72-93	construct of wild type Ltn consisting of residues 72-93
zIc	internal fragment with ‘z’ fragment at their N terminus and a ‘c’ fragment at their C terminus
Zip	leucine zipper motif

Physical constants and quantities

B	magnetic field strength
d	distance
e	elementary charge (= 1.60217585 x10^{-19} C)
E	electric field strength
E_K	kinetic energy
f	frequency
K	mobility
K_a	association constant
k_b	Boltzmann constant (= 1.3806488(13) x10^{-23} JK^{-1})
K_d	dissociation constant
L	length
m	mass
N	number density
n	stoichiometry of binding
P	pressure
r	radius
R_g	radius of buffer gas
T	temperature
t	time
U	DC potential
V	peak amplitude of RF potential: electrical potential difference
V	voltage
v	velocity
z	nominal charge
ΔG^0	Gibbs free energy
ΔH^0	enthalpy
ΔS^0	entropy
Φ_0	total electric potential
μ	reduced mass
ρ	density
ω	angular velocity
Ω	momentum transfer integral (collision cross section)

1

Introduction

In the field of protein science it was, until recently, generally accepted that a protein will possess a single unique, evolutionary constrained three-dimensional structure. However, numerous cases have demonstrated that proteins and peptides undergo conformational changes upon binding, display conformational flexibility or even adopt multiple distinct conformations in equilibrium. In order to study conformational plasticity in dynamic systems, techniques must be capable of considering multiple conformations. Mass spectrometry and ion mobility-mass spectrometry, in combination with other biophysical techniques, can provide invaluable insight into such systems.

1.1 Protein structure

Proteins are fundamental molecules of life which carry out numerous essential functions; providing structure, mediating the immune response, catalysing reactions and transporting small molecules, to name only a few[1]. The basic building blocks of proteins are amino acids, there are 20 common amino acids which differ only with respect to their side chain, the structures of which are shown in Appendix 1 Table A1.1. In proteins and peptides, amino acids form long sequences in which the peptide bond is most frequently planar and *trans* and the sequence of amino acids encodes the three-dimensional structure of the protein[2,3].

The secondary structure is the local structure adopted by amino acids in the protein, which is stabilised through hydrogen bonds between the amino- and carboxy-groups of the peptide bonds. Common secondary structural motifs include the α-helix, β-strand, turns and coils[4]. The α-helix is a right handed coil, in which each turn is ~5.4 Å and comprises 3.6 amino acids which orientate themselves with the side chains pointing outwards[5,6]. The amino acids alanine, leucine, methionine, lysine and uncharged glutamate have the highest α-helix-forming propensity[7]. The β-strand motif is an extended stretch of amino acids, typically 3-10 amino acids, with no hydrogen bonding occurring between carboxy and amino groups in neighbouring amino acids, however, strands can lay next to each other, in either a parallel or anti-parallel arrangement promoting inter-strand hydrogen bonding and stabilisation[3]. In addition regions of no defined structure can be present such as random coil and intrinsically disordered regions[8], composed of high numbers of polar residues including arginine, glycine, glutamine, glutamic acid, serine, proline and lysine[8-10]. Secondary structural elements can then be arranged into super secondary structural motifs including helix bundles, coiled-coils, β-hairpins, β-barrels or 'jelly-roll' topologies[3]. The tertiary or global fold of the protein describes all secondary and super secondary structural motifs and their interactions.

For over a century a central tenant in structural biology has been that the function of a protein depends upon its well-defined three-dimensional structure[11]. Through the discovery of functional disordered proteins[12,13] and proteins which can undergo conformational switching[14-16] it is apparent that it is necessary to consider the dynamic nature of proteins, the possibility of conformational switching and conformational ensembles in conjunction with function and therefore studies have to consider conformational flexibility.

The focus of this thesis is to study both the structure and interactions of biologically relevant proteins and peptides, using techniques capable of the consideration of multiple species and

conformations existing in equilibrium. Here a focus is placed on mass spectrometry (MS) and ion mobility mass spectrometry (IM-MS) techniques, assessing the information which can be obtained using these experimental approaches. The aim of the studies presented is to demonstrate the detail which can be gleaned from these techniques in addition to their application to ongoing research challenges. The following sections introduce the systems under study considering both their structure and functions (section 1.2-1.4), and the basis of the experimental techniques applied (section 1.5).

1.2 MYC

1.2.1 Classification and structure

The MYC family of transcription factors contains four members; c-MYC, N-MYC, L-MYC and the lesser studied S-MYC. MYC proteins have been described as possessing a number of conserved domains including an N-terminal transactivation domain (TAD), regions known as MYC boxes which play important roles in binding to co-activators and transcriptional activation, a central region and a C-terminal domain (CTD) essential for DNA binding[17]. Additionally c-MYC contains a region known as the nuclear localisation signal (NLS), see Figure 1.1. The CTD contains a basic helix-loop-helix leucine zipper (b-HLH-Zip) domain which is capable of binding a similar region in the specific binding partner of MYC, MAX[18,19], forming a two stranded parallel coiled-coil (Figure 1.1).

The b-HLH-Zip motif is particularly common in DNA binding proteins[20] and can be thought of as two connected regions, firstly the basic helix-loop-helix region characterised by two α-helices connected by a loop, which is thought to be important for binding to DNA. The second region, the leucine zipper forms an extended coiled-coil, with each coil comprising a conserved leucine residue at every seventh position. These amino acid sequences fold into an α-helical arrangement where the leucine residue side chains extend out from the helix and form non-covalent interactions with a corresponding leucine residue on a matching helix in a second polypeptide chain (Figure 1.1). The leucine zipper is further stabilised by electrostatic interactions and inter-helical salt bridges, which are thought to confer specificity to its formation[21,22]. These regions are disordered until they bind, forming the two stranded parallel α-helical coiled-coil, and are thus an example of coupled folding and binding. It is only after interaction with its binding partner that c-MYC is then capable of carrying out its function as a transcription factor (see section 1.2.2).

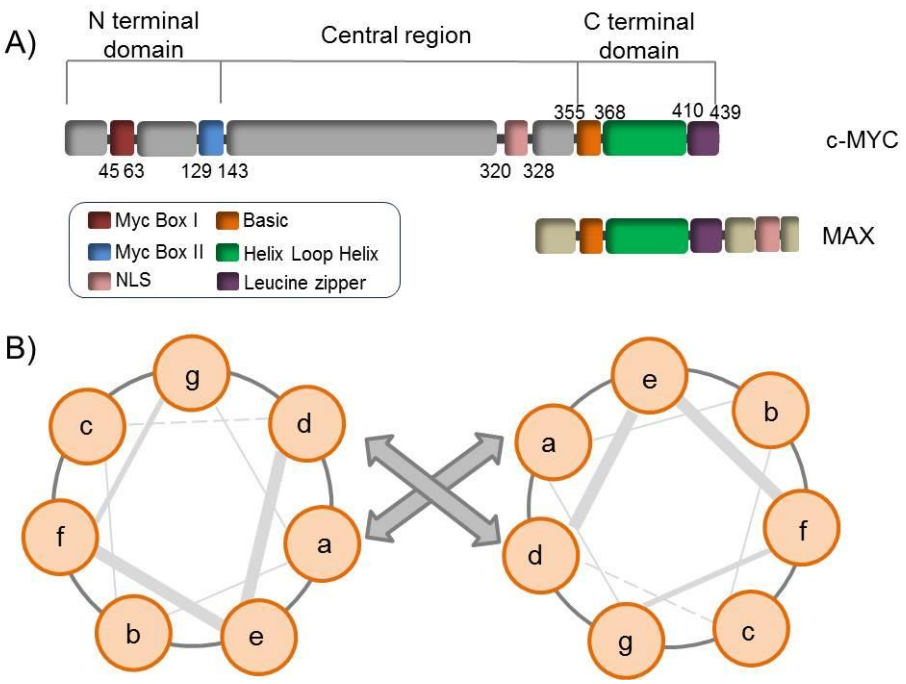

Figure 1.1: A) domain structure of c-MYC and MAX, B) helical wheel representation of the leucine zipper domain of c-MYC:MAX annotated with the classical abcdef heptad, solid arrows represent hydrophobic interactions, modified from reference[19].

Interestingly, c-MYC and MAX have also been found to exist as a antiparallel tetramer in the crystal structure, which, due to the low dissociation constant, is believed to also exist *in vivo*, however, the function of such a complex remains elusive[23].

1.2.2 Function

The MYC family of proteins are transcription factors, proteins which bind to specific DNA sequences controlling information transfer from DNA to messenger RNA, regulating cell development, differentiation and cell growth[24].

In order to carry out these functions MYC has to bind its binding partner MAX; it is through the formation of MYC:MAX heterodimers that MYC proteins can bind to DNA and activate transcription[25,26]. Interactions between the b-HLH-LZ domain of MYC and proteins other than MAX are also thought to contribute to biological function, to a lesser extent, however, monomeric MYC species are inactive[25]. MYC can both activate and repress transcription depending on the binding partner and cellular environment[25]. Recent reports have suggested that MYC can bind to ~25,000 sites in the human genome[27-29] and hence MYC proteins are thought to regulate ~10-15 % of all genes through recruitment of co-factors[30].

MYC expression is regulated through a number of external signals, including growth factors, extracellular matrix contacts and through internal cell cycle control[31], and can be deregulated through several mechanisms which act directly or indirectly to target its expression and/or activity[32]. Modifications in MYC expression can result in alterations to the cell apoptotic pathway , uncontrolled cell proliferation, transformation or genomic instability[17]. The MYC proteins have hence been implicated in a wide range of aggressive human cancers from haematological malignancies to solid tumors and often have poor prognosis[17,33]. Oncogenic alterations of MYC are induced by point mutations, gene amplification and translocation, over-expression and enhanced translation[17], however, the majority of cases show no distinguishing mutations[34,35].

Due to the high number of different human cancers deregulated MYC is implicated in, and the increase in rational drug design targeting proteins involved in regulation of cell proliferation and/or apoptosis, MYC has become a target for novel cancer therapies. There are a number of reported approaches through which MYC can be targeted at different levels that can be employed either on their own or in combination with current cancer therapies. Strategies to-date include inhibition of *myc* gene expression at DNA and mRNA levels through interaction with the *myc* promoter and interference with transcription, inhibition of MYC expression, disruption of MYC:MAX dimerization, inhibition of the interaction of MYC with other binding partners, interference of MYC:MAX binding to DNA, inhibition of expression of MYC target genes, and promoting MYC protein degradation[17,36,37].

1.2.3 c-MYC and its implication in cancer

Of the MYC proteins the founding member c-MYC has been implicated in cancer to the greatest extent and hence is the most widely studied. Furthermore, mouse models have shown that inactivation of c-MYC can result in tumour shrinkage with minimal side effects, hence c-MYC inhibitors are potential attractive targets for chemotherapy[38,39]. Targeting c-MYC with the aim of disrupting formation of its heterodimeric complex with MAX is proving particularly attractive and a number of small molecule inhibitors have been reported[40-42], to-date, however, no anti-MYC drugs exist on the market.

c-MYC is an interesting target for study from a structural point of view due to its conformational plasticity. Furthermore a fast, efficient and reliable method to probe the interaction and inhibition of c-MYC:MAX in the presence of potential drug candidates is vital to aid rational and intelligent design of new inhibitors, therefore, from a functional

view, c-MYC is an essential target for study. Chapter 3 assesses the potential of mass spectrometry and ion mobility mass spectrometry to study c-MYC:MAX.

1.3 Chemokines

1.3.1 Classification and structure

Chemokines are a large subset of cytokines[43]. Cytokines are a group of signalling proteins which derive their name from the Greek words for cell (*cyto*) and movement (*kinos*). Chemokines (~8-14 kDa) are so called due to their ability to induce chemotaxis (chemically promoted movement). Chemokines generally contain two disulfide bonds and have been subdivided into four subclasses based upon the pattern of conserved cysteine residues near the N-terminus[44]. The CC, CXC and CX_3C subclasses contain two cysteines at the N-terminus either side by side, separated by one or three amino acids, respectively. The final subclass is the C subclass and is unique in that it contains a single disulfide bond. The genes of the different chemokine subclasses have been found clustered in certain chromosomal locations. In humans CC chemokines are clustered on chromosome 17, CXC chemokines on chromosome 4, the gene of the sole member of the CX_3C subfamily is found at chromosome 16 and the two variants of the C subfamily are found at chromosome 1[45]. Additional minor clusters have also been found for CC and CXC subfamilies.

Chemokines display low amino acid sequence homology, however, their three-dimensional structure is highly conserved, as shown by nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography[46-48]. The tertiary fold comprises an extended N-terminal loop (important for receptor binding), a central three stranded antiparallel β-sheet (providing a stable core), and a C-terminal helix which is thought to confer stability to the overall structure by folding over the β-sheet[49,50]. Chemokines have been reported in monomeric, dimeric and tetrameric forms. Notably, the dimers display conserved dimerisation motifs which depend upon the family to which they belong, either CC or CXC[51]. The monomeric form of the protein is, however, thought to be the active form for receptor binding[50]. Chemokines can be found in mammals, fish, birds and even some strains of virus and bacteria[52]. The work presented in this thesis considers human chemokines and in particular the C chemokine lymphotactin (section 1.3.3).

1.3.2 Function

Chemokines and chemokine receptors have been referred to as the *'eyes and ears of the immune system'*[53], guiding cells of both the innate and adaptive immune system to specific locations, in addition to performing other cellular migratory events involved in organ development, wound healing and angiogenesis[54,55]. Chemokines can be broadly classified into inflammatory chemokines, which are expressed in response to inflammatory stimuli, or homeostatic chemokines which are expressed in the absence of inflammatory stimuli or by cells which do not typically act as inflammation inducers, however, some chemokines can act as both[56,57].

Chemokines are best known for their role in immune response, where they control the migration and activation of leukocytes[50,58]. Chemokines function by activating specific G protein-coupled receptors (GPCRs), resulting in signal transduction and the migration of inflammatory and non-inflammatory cells to appropriate tissues or compartments[49]. Cellular migration results as a consequence of several signalling events and intracellular changes, for example cell shape change, actin polymerisation and receptor polarization have been implicated[59]. In the case of leukocyte migration in response to inflammation the process involves three stages; vascular adhesion, extravasation and chemotaxis. In vascular adhesion, leukocytes interact with adhesion molecules on the luminal side of blood vessels[60]. Integrin function is up-regulated *via* chemokine receptor signalling, which results in firm attachment of leukocytes to the blood vessel wall. This firm attachment enables leukocytes to migrate across the endothelium and, once in the perivascular area, leukocytes are orientated and directed through a chemokine gradient. Interactions with glycosaminoglycans are thought to play an important role in the localisation of chemokines and development of the concentration gradient, perhaps as a method to overcome the effects of vascular flow[61-63]. This will be considered in Chapter 5 of this thesis.

There are ~50 human chemokines, which can activate 19 different transmembrane receptors[50]. Many chemokines of the same subfamily bind the same receptor, suggesting there is a redundancy to this system[45,64]. It is thought, however, that response specificity may be due to expression patterns of individual chemokines or that different combinations of activated receptors yields different biological responses[45]. Inappropriate regulation of chemokines is associated with a number of pathologies including inflammatory diseases, cancer, HIV-associated disease and cardiovascular disease[65,66].

1.3.3 Lymphotactin

Lymphotactin (Ltn) is a unique chemokine for many reasons. It contains only a single disulfide bond and hence is the defining member of the C subclass of chemokines[67], additionally Ltn contains a unique extended C-terminal region and most significantly Ltn is a metamorphic protein[68]. Metamorphic proteins can exist in multiple distinct conformations. Human Ltn has been shown to adopt two distinct folds, a monomeric and dimeric fold, in equilibrium under physiological conditions. The Ltn conformational equilibrium, however, is sensitive to changes in ionic strength and temperature which has enabled representative structures of both folds to be solved[69-71].

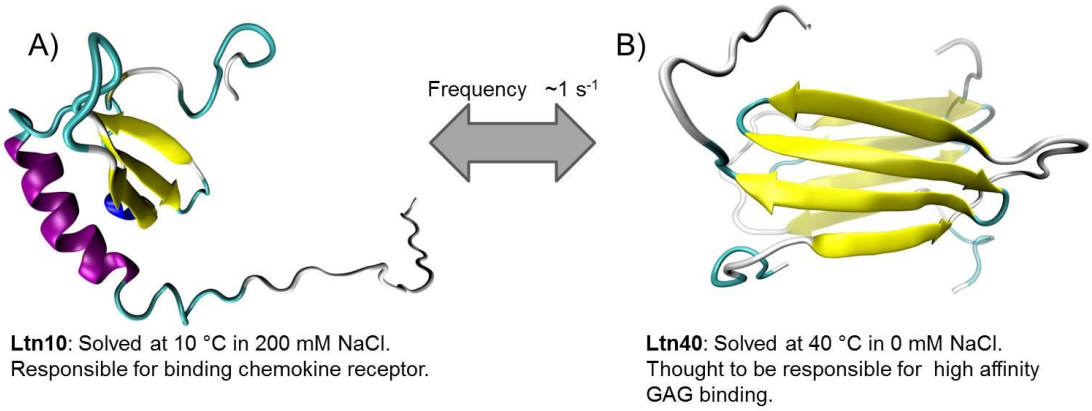

Figure 1.2: Ltn conformations, A) Ltn10 (PDB 1J9O) and B) Ltn40 (PDB 2JP1).

The first fold to be solved was the conserved chemokine conformation consisting of three β-strands and an α-helix with the extended C-terminal sequence forming an intrinsically disordered (ID) tail. This fold is often referred to as Ltn10 as it was solved at 10 °C (Figure 1.2). Ltn10 is thought to be responsible for activating the G protein coupled receptor XCR1. Interestingly, the sequence of the ID tail is similar across species, with human and mouse Ltn exhibiting 68 % sequence identity in the final 25 amino acids, compared with 58 % overall sequence identity[72,73]. The conservation of this intrinsically disordered tail implies functional importance, however, it has been shown to be completely dispensable for XCR1 activation[74] and to-date its function remains unknown. The second distinct fold of Ltn (Ltn40) is a novel dimer which is formed from two identical units comprising four β-strands and the ID tail (Figure 1.2). This form is thought to be responsible for high-affinity glycosaminoglycan binding and hence both forms are believed to be functionally essential *in vivo*[71]. Interconversion between the folds is reversible and involves complete restructuring of all core residues, occuring on the time scale of seconds[75].

Lymphotactin homologs have been found in many species including humans, mice, rats and chicken[76]. Ltn (also known as XCL1) can be expressed in many tissues and has been found in thymus, spleen, small intestine and peripheral blood lymphocytes and at lower levels in the lung, colon, ovary and testis[72,73,77]. Ltn is produced through T-cell receptor activation in $CD4^+$ and $CD8^+$ T-cells and to a lesser extent through natural killer and γδ T cells[67,78,79], and is known to bind XCR1 to induce T-cell and NK cell chemotaxis[67]. A gene encoding a highly related chemokine known as lymphotactin-β or XCL2 has been found in humans which differs by only two amino acids at positions seven and eight[80]. The significance of XCL2 and extent to which it is produced *in vivo* remains unknown; therefore, structural and functional investigations have focused on XCL1.

Significantly, Ltn has been used in the development of novel alternative cancer therapies[81-83] and Ltn has been implicated in a number T-cell-mediated disease states[68]. Further structural and functional studies of this protein are therefore required as future development of inhibitors or mimics of Ltn may have therapeutic potential. Human Ltn is an excellent example of structural plasticity in proteins and makes an interesting target for evaluation of new approaches to consider multiple dynamic conformations of proteins. In Chapter 4 the applicability of MS and IM-MS based methods to study Ltn are assessed, and their potential to study future mimics or inhibitors considered. Chapter 5 focuses on the use of MS, IM-MS to study the interactions of this novel chemokine with other biomolecules, namely glycosaminoglycans.

1.4 Human β-defensins

1.4.1 Classification and structure

Defensins are a class of small (2-6 kDa) cysteine-rich peptides, which are part of the large family of cationic host defence peptides[84]. Defensins are found across species and have been reported in mammals[85], insects[86] and plants[87]. Mammalian defensins contain six conserved cysteine residues which form three disulfide bonds, the connectivites of which can differ and hence these peptides are subdivided into three subfamilies (α, β and θ) based upon the position and connectivity of the cysteine residues[88]. In α-defensins the connectivities are Cys^{I}-Cys^{VI}, Cys^{II}-Cys^{IV} and Cys^{III}-Cys^{V}, whilst in β-defensins the connectivity is; Cys^{1}-Cys^{V}, Cys^{II}-Cys^{IV} and Cys^{III}-Cys^{VI}. The θ-defensins are cyclic peptides which possess the same connectivity as α-defensins, and have been found in a number of primates but not in

humans[89,90]. Despite their differences, the defensin subfamilies are thought to share a common ancestry[91], with the β-defensins being the ancestral gene.

Genomic studies have predicted the sequences of over 30 human β-defensins (HBDs)[92], however, to-date only the native peptide sequences and structures have been studied for HBD1, HBD2 and HBD3. Despite a low sequence homology, X-ray crystallography and NMR studies of HBD1-3 display notably similar tertiary folds. These comprise an N-terminal α-helix and three anti parallel β-strands stabilised by the three conserved disulfide bonds[93-96], which is strikingly similar to the tertiary fold of β-defensins from different species[97,98]. Studies have demonstrated that the majority of β-defensins exist as monomers in solution[94,96,97], however, the crystal structure of HBD2 shows evidence for dimer and octamer formation[93]. Furthermore, HBD3 is thought to be a dimer in solution[96]; however, the structure is yet to be solved.

1.4.2 Function

Mammalian β-defensins are predominately expressed in epithelial cells and leukocytes[99]. The first human β-defensin (HBD1) was identified from purification of hemofiltrate[100] but has subsequently been found to be expressed throughout epithelial cells in the human body[101]. HBD1 is expressed constitutively and is thought to act as an initial line of defence against microbial infection. HBD2 and HBD3 were first isolated from psoriatic scales[102,103], and have since been found in numerous other tissues including the airway and oral tissue[104]. HBD2 has additionally been found in gastrointestinal, female reproductive tract, kidney, ear and eye tissues[104]. Both HBD2 and HBD3 expression is induced upon bacterial challenge and inflammation[105].

β-defensins are thought to play a role in innate and adaptive immunity[88]. The innate immune system is characterised by a series of non-specific mechanisms to overcome infection, and it is through such interactions HBDs exhibit a wide range of antibacterial, antifungal and antiviral activities[106,107]. HBD1 and HBD2 are potent towards Gram-negative bacteria and are only weakly active against Gram-positive bacteria[102] whereas HBD3 has been shown to be active against both[103]. Studies have shown that HBD1 has only minor antibiotic activity, however, upon reduction of its disulfide bonds, potentially due to increased conformational freedom and a corresponding increase in available surface charges, it exhibits potent antimicrobial activity against the fungus *Candida albicans*, whilst under the same conditions activity of HBD3 was not improved[108,109]. Structural studies have also shown that incorrect connectivity of the cysteine residues in HBD3 does not diminish its antimicrobial activity

against *E. coli* and a similar observation was made for mouse β-defensin 14[110,111]. Consideration of the fold and flexibility with respect to function is therefore important for these antimicrobial peptides.

The mechanism by which defensins kill bacteria is complex and still under investigation. The first stage is thought to involve accumulation of the peptides at the microbe surface, facilitated through electrostatic interactions between these highly basic peptides and negatively-charged components of the outer bacterial cell wall[112]. The antimicrobial activities of both HBD1 and HBD2 are reduced by physiological salt concentrations, however, HBD3 antimicrobial activity is not affected by salt concentrations[103,113,114]. Salt can act to weaken the electrostatic attraction to the membrane, and hence the high net charge of HBD3 is thought to overcome the detrimental effects of salt.

Most antimicrobial peptides kill bacteria by attacking their cell membrane and causing cell lysis, however, some antimicrobial peptides have been shown to permeate through the membrane and act upon specific targets[115]. There are two widely accepted methods by which antimicrobial peptides can cause lysis; the carpet and pore models[112,116-118]. In the carpet model peptides cover the membrane, structurally weakening it and at a critical concentration the membrane is broken. In the pore model peptides transverse the membrane, to form a pore through which the cell cytoplasm can be lost. To-date, however, the mechanism through which β-defensins exert their activity has remained elusive[85]. Studies of HBD3 and S. *aureus* suggested that cell lysis was not the dominant mechanism and instead interference with the cell wall synthesis was thought to prevail[119,120]. Shorter peptide model systems of HBD1-3 suggest that the peptides can interact with negatively charged lipid head groups causing membrane destabilisation, which can affect the efficient functioning of cytoplasmic membrane proteins in bacteria and results in cell death[121]. The mode of action of HBDs hence remains an active area of research.

More recently β-defensins have been found to display chemotactic behaviour, recruiting immune cells to sites of infection through a peptide concentration gradient, and hence play a role in adaptive immunity[99,122]. Human β-defensins have been found to be capable of recruiting both T-cells and immature dendritic cells, thought to occur through interaction between the β-defensin and chemokine receptor-6 (CCR6)[122]. HBD3 has also been shown to interact with CCR2 on myeloid cells which can also result in chemoattraction[123,124]. As with chemokines, HBD2 has been shown to interact with glycosaminoglycans, an interaction which may also prove important with respect to chemotaxis[125]. Interactions with chemokine receptors and chemotactic activity are clearly important immunoregulatory functions for

β-defensins and hence are areas of on-going research. Structural studies have shown that the connectivity of the six conserved cysteine residues, and the resulting folds, can have significant effects on the chemotactic activity of HBD3 through CCR6[110]. More recent studies have also highlighted that Cys^{v} and isoleucine/leuince at the N-terminus are essential for chemokine activity[126,127]. Consideration of sequence and fold is therefore once again essential for understanding the biological activities of these peptides.

Converse to their roles in the innate and adaptive immune system, HBDs have also been implicated in disease, which has been described as '*the yin and yang of human β-defensins'*[128]. Detrimental effects include promotion of tumour growth following over-expression of HBD3[123,129]. High levels of HBDs have also been found in psoriasis lesions[102,130,131], however, these high concentrations are thought to provide a level of protection against incoming pathogens. Sufferers of the inflammatory skin disease atopic dermatitis frequently experience infection of lesions which has been associated with low concentrations of β-defensins[132].

In order to better understand the myriad of functions of β-defensins, it is instructive to consider not only their interactions but also conformational effects upon binding. The work presented in Chapter 6 of this thesis aims to characterise the interactions of β-defensins with glycosaminoglycans using biophysical techniques, including isothermal titration calorimetry and transmission electron microscopy in combination with mass spectrometry and ion mobility mass spectrometry.

1.5 Biophysical techniques for studying protein and peptide structure and interactions

1.5.1 High resolution techniques for structure determination

X-ray crystallography and NMR spectroscopy are the highest resolution techniques available for the study of protein structure, both of which enable detailed models of the three dimensional fold to be built. NMR and X-ray crystallography have undoubtedly provided some remarkable results[133-137] with unparalleled resolution, however, these techniques often suffer due to the sheer amount of material needed, an inability to produce to high resolution crystals and protein aggregation. Furthermore, through the recent discovery of a number of proteins which exist in equilibrium between multiple conformations in a ligand-free state it

has been suggested that traditional techniques have been inadvertently selecting against the consideration of proteins that can adopt multiple conformations[14].

1.5.2 Relevance of gas-phase techniques in the study of proteins and peptides

Recently gas-phase techniques, including mass spectrometry (MS), ion mobility-mass spectrometry (IM-MS) and gas-phase infrared spectroscopy (IR) have been applied to study the structure and interactions of peptides and proteins. A central question regarding the use of gas-phase techniques to study biological molecules is to what extent, if any, these molecules can retain their solution or native-like folds in the gas-phase.

It has been previously demonstrated that the mass spectrum obtained for a protein from electrospray ionisation (ESI) or nano-electrospray ionisation (n-ESI) is highly dependent upon the solution conditions and solution state of the protein prior to ionisation and desolvation[138-141]. The observed charge states of proteins and peptides depend upon the availability of ionisable sites, with multiply charged ions forming due to charge accumulation in the droplets. Spectra acquired from buffered 'native-like' conditions generally present lower charge states and with narrower charge state distributions than spectra acquired from solutions containing a high proportion of organic solvents and/or lower pH (denaturing conditions) which gives rise to higher charge states and a wider distribution of charges than the native-like species, partially due to more unfolded species having more available chargeable sites.

Gas-phase studies have provided evidence for preservation of secondary structural motifs in proteins and peptides in the gas-phase, including α-helices[142-144] and to a lesser extent β-sheets which have mainly been studied in small model peptides[145,146]. Retention of tertiary folds of proteins has been supported by ion mobility-mass spectrometry (IM-MS) coupled with molecular dynamics simulations[147] (section 1.5.4). In solution, protein folds can be stabilised through non-covalent interactions including hydrogen bonding, electrostatic interactions, van der Waals forces and hydrophobic interactions. A report presented by Breuker and McLafferty considered to what extent these interactions can be preserved in the gas-phase[148], and highlighted that charged side chains can collapse on the order of picoseconds, hydrophobic interactions and dissociation of electrostatic interactions can occur over milliseconds, electrostatic interactions such as salt bridges can also strengthen or be formed in the solvent-free environment, following which new non-covalent interactions can also be formed over seconds to minutes. Therefore the extent that protein fold is preserved in

the gas-phase depends in part upon the time ions spend in the gas-phase, which is hence a key consideration for structural studies using these techniques. Gas-phase techniques can be advantageous, however, as they provide an environment in which intrinsic and extrinsic factors governing protein fold and stability can be distinguished and studied.

The soft ionisation techniques of ESI and n-ESI (see section 1.5.3.1) enable preservation of non-covalent complexes[149] including: protein:ligand complexes[150-152], protein:protein complexes[153,154], even very large complexes such as virus capsids[155], as well as protein or peptide aggregates particularly of amyloid-forming species[156-158], with comparison to other techniques providing evidence that solution states can be preserved. Another highly influential study was carried out in the laboratory of Graham Cooks and showed that the enzymes lysozyme and trypsin could be transmitted through a mass spectrometer before being collected by soft-landing on a surface, following which they retained their biological function[159].

Finally, it could also be argued that the gas-phase environment of the mass spectrometer where the dielectric constant is 1, is more similar to that which proteins and peptides experience in their cellular environment (2-6) than in a dilute aqueous solution[160]. Hence MS and IM-MS have promise for the study of biological molecules, enabling consideration of isolated molecules. A particularly attractive feature of these gas-phase techniques is that successful transmission of biological molecules does not depend on retention or preservation of a single constrained fold. Following gentle desolvation, multiple different folded, unfolded and partially unfolded species may all be present and selectively studied, enabling dynamic conformations and protein unfolding to be studied. Hence these techniques are particularly advantageous for the study of dynamic proteins.

1.5.3 Biological mass spectrometry

The foundations for mass spectrometry were laid 100 years ago, through the pioneering work of Thomson and Aston in the development of the mass spectrograph, as a method to study gas-phase atomic structure[161-163]. Since its development and the countless improvements made since then, MS has found widespread use across many scientific fields, including the analysis of proteins and peptides[164,165].

In brief, MS is a technique through which the mass of molecules can be determined, based upon the motion of a charged particle in an electric field, which in turn is centred upon the molecules mass to charge ratio (*m/z*). There exist many different types of mass spectrometer, however, in all cases similar functions are performed: firstly gaseous ions are produced and

introduced into a region of high vacuum where they are subjected to magnetic and/or electric fields; next ions are sorted according to their *m/z* and the abundance of each *m/z* is measured[166]. This thesis employs the use of both quadrupole time of flight mass spectrometers (Q-TOF) and a Fourier transform ion cyclotron mass spectrometer (FT ICR-MS), hence the following sections will discuss only these types of instruments (see section 1.5.3.2). Detail on other instrumentation and further information on those discussed can be found elsewhere[167,168].

1.5.3.1 Ionisation methods

The first stage of any mass spectrometry experiment involves ionising and introducing the sample into the mass spectrometer. The development of soft ionisation techniques such as matrix-assisted laser desorption ionisation (MALDI) [169,170] and electrospray ionisation (ESI) were major breakthroughs enabling the analysis of complex biomolecules such as proteins and peptides by MS. In MALDI species are generally co-crystallised with organic acids, resulting in low pHs hence for all studies reported in this thesis ESI or n-ESI were used.

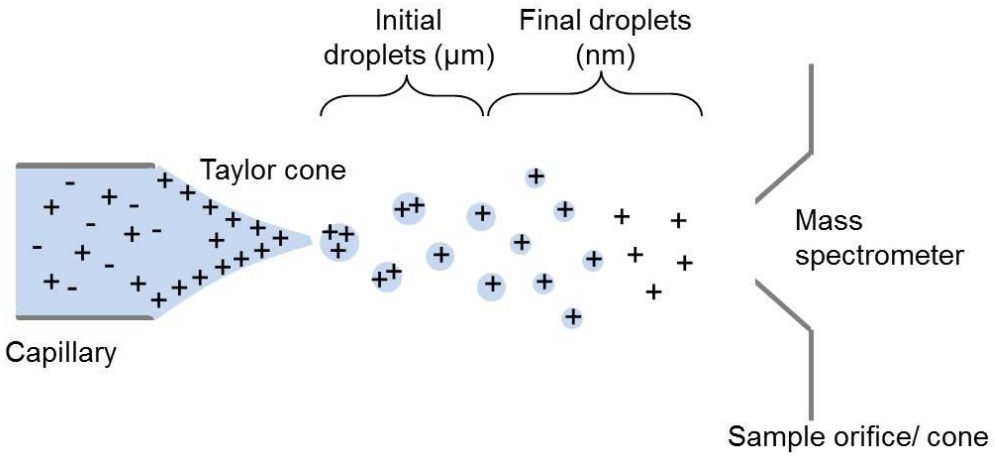

Figure 1.3: Schematic representation of the ESI in positive mode.

The origins of electrospray ionisation lie with Dole *et al*[171-173] who reported its use as a method to generate *in vacuo* beams of macroions. Fenn *et al*[174,175] further developed this technique and revolutionised mass spectrometry by enabling the transfer of large biomolecules into the gas-phase as multiply charged ions, permitting their molecular weights to be determined and *'giving wings to molecular elephants'*[176]. In ESI the sample solution is passed through a capillary at atmospheric pressure to which a high electric potential is applied. ESI can produce either positive or negative ions depending on the potential applied, generally in the range ±2-6 kV. In this thesis ionisation was performed in positive mode in all cases and hence discussion here will focus on this mode. The potential difference between

the sample capillary and the instrument inlet creates an electric field, causing an accumulation of positive charges at the liquid surface at the end of capillary, which then breaks to form highly charged droplets. The solution at the tip of the capillary is distorted into a Taylor cone, from which a fine mist of droplets is produced[177,178], assisted by a coaxial flow of nitrogen. The droplets emitted experience rapid solvent evaporation, aided by a heated source, which causes an increase in charge density until the Rayleigh limit[179] is reached, when surface tension is balanced by Coulombic repulsion. At the Rayleigh limit, droplets undergo fission producing smaller charged species. Repeated evaporation and fission events yield desolvated ions which are then accelerated into the mass spectrometer (Figure 1.3).

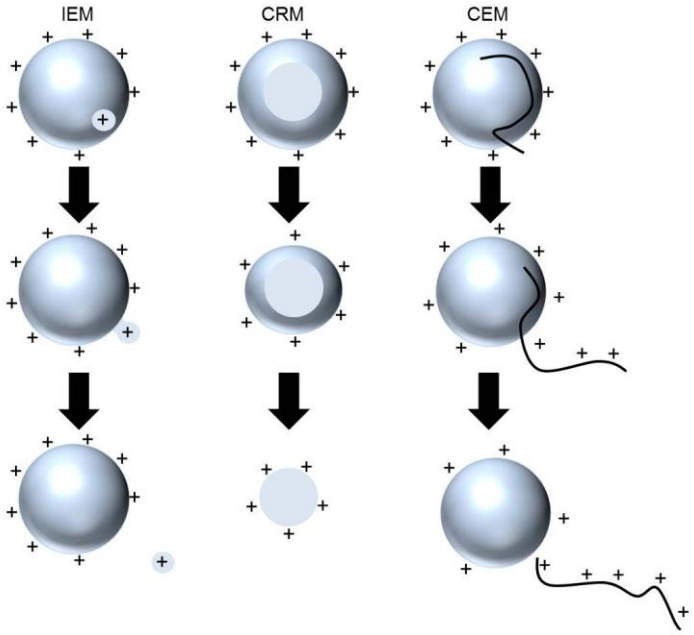

Figure 1.4: Summary of ESI mechanisms; ion ejection model (IEM), charge residue model (CRM) and chain ejection model (CEM), modified from reference[180].

The mechanism occurring in the final stages of gaseous ion formation in ESI is still under debate; however, there exist three widely accepted models for the process (Figure 1.4)[180,181]. The ion evaporation model was described by Iribarne and Thomson[182] and proposes that the electric field in a Rayleigh-charged nanodroplet would be suitably high to cause ejection of a charged analyte ion; this model is suggested to be most appropriate for ions of low molecular weight. The charged residue model proposed by Dole *et al*[171] suggests that the Rayleigh-charged nanodroplets which contain a single analyte evaporate to dryness, with the charge present on the last solvent shell being transferred to the analyte upon evaporation. The charged residue model is widely accepted for large globular species, such as natively folded proteins[183,184]. Finally, the chain ejection model has been recently presented for disordered

polymers and unfolded proteins[185,186]. Here, the hydrophobic nature of the extended species makes residing within the interior of the Rayleigh-charge droplet unfavourable causing migration to the droplet surface where one terminus gets ejected into the vapour phase followed by sequential ejection of the extended chain, leading to high charge states for these species.

Advances in ESI saw the development of a low flow rate version known as nano-electrospray ionisation (n-ESI), which has the advantage of lower sample rates, hence lower volumes and also affords the use of lower concentrations[187,188]. n-ESI is particularly advantageous in biological studies where sample is often limited. In n-ESI samples are ionised from fine tipped glass capillaries with an internal diameter of ~0.1 mm, to which a voltage of ±1-2 kV is applied either *via* a metal coating on the capillary or through an inert metal wire placed within the capillary. The processes occurring in ESI and n-ESI are thought to be the same; however, in n-ESI the droplets formed are smaller, resulting in a more stable spray[188,189].

1.5.3.2 Mass analysers

There are many different mass analysers which can be employed individually or in combination in hybrid mass spectrometers[167].

Quadruple mass analyser

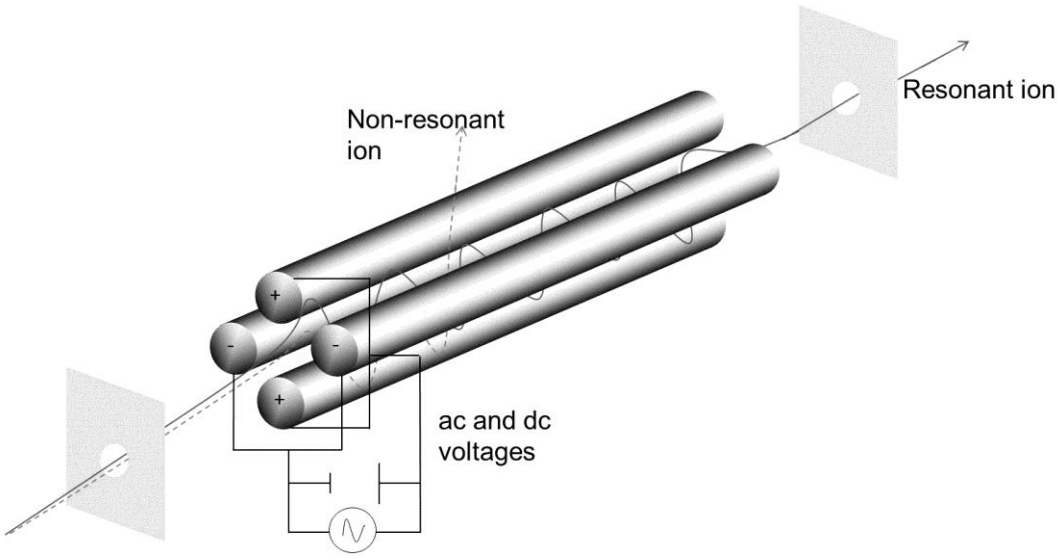

Figure 1.5: Schematic representation of the quadrupole analyser.

A quadrupole mass analyser consists of four rods of circular or hyperbolic section, which are arranged in pairs (Figure 1.5). A fixed direct current (DC) and an alternating radio frequency

(RF) are applied to the pairs of rods, so that when one pair of rods is in-phase (*e.g.* positive) the other pair is out of phase (*e.g.* negative) with respect to RF potential, thus creating an oscillating electric field as described by Equation 1.1[167].

$$\pm\Phi_0 = \pm(\mathrm{U} - \mathrm{V}\cos\omega t) \qquad [1.1]$$

Where, Φ_0 is the potential applied to the rods, U is the DC voltage, V the zero-peak amplitude of the RF potential, ω the angular frequency of the RF potential and *t* is time.

An ion entering the quadrupole analyser will be attracted to a rod of opposite polarity, however, as an alternating RF potential is applied to the pairs of rods their polarity is constantly changing and hence if a rod changes polarity before the ion collides with it the ions trajectory will change. Ions therefore travel in an oscillating path (Figure 1.5), which is dependent on the values of U, V and ω (equal to 2π*v*, where *v* is the RF frequency). If the RF amplitude (V) is scanned, ions of different *m/z* can be transmitted through the length of the analyser, which is most frequently the mode of operation in MS experiments. For an ion of specific *m/z*, however, there are specific values of U, V and ω which provide a stable trajectory and, hence, if these variables are fixed only the ion of interest will pass through the quadrupole and all ions of differing *m/z* will collide with the rods. The quadrupole is operated in this mode to act as a mass filter, prior to fragmentation studies, where ions of a particular *m/z* will be transmitted into a collision cell.

Time-of-flight analyser

In Q-TOF instruments, after passing thorough the quadrupole and other lens present, ions pass into the time-of-flight (TOF) analyser. In its original linear arrangement packets of ions enter the analyser where they are accelerated into a field-free flight tube by a potential difference. Once in the field-free region their velocity is dependent on their *m/z*, and hence the time taken for an ion to travel to the detector can be used to determine the *m/z* of the ion, as shown in Equations 1.2-1.4.

$$E_K = \frac{mv^2}{2} = zeV \qquad [1.2]$$

Where, E_K is the kinetic energy of the ion, *m* is the mass, *v* is the velocity of the ion entering the TOF, z*e* is the total charge of the ion and *V* is the accelerator voltage.

$$t = \frac{d}{v} \qquad [1.3]$$

$$t^2 = \frac{m}{z}\left(\frac{d^2}{2eV}\right) \qquad [1.4]$$

The time taken (t) for an ion to travel the length of the drift tube (d) is given by Equation 1.3, which when combined with Equation 1.2, demonstrates how the *m/z* is related to the time taken for an ion to traverse the field-free region (Equation 1.4).

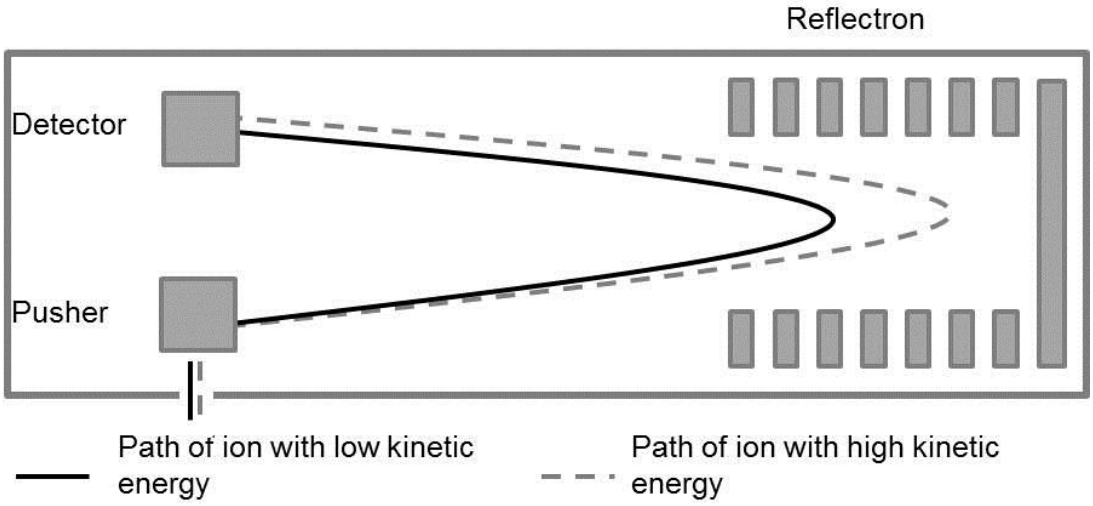

Figure 1.6: Schematic representation of the time-of-flight (TOF) analyser.

The resolution of linear TOFs are limited due to a number of factors including length of the drift region, variation in the initial kinetic energy of the ions as well as time and spacial distribution effects. In 1973 Mamyrin *et al*[190] proposed the use of the reflectron to overcome the loss of resolution due to variations in kinetic energy. A reflectron is a series of ring electrodes and grids which act as an ion mirror, reflecting ions back up the flight tube. The reflectron TOF increases resolution by enabling small differences in kinetic energy to be corrected, whilst also increasing the effective drift length. A TOF is frequently coupled to the back end of a quadrupole analyser, creating a quadrupole time-of-flight (Q-TOF) mass spectrometer. Commercially available Q-TOFs can have a high mass range, which can be extended further through modifications[191]. The coupling of Q-TOFs with soft methods of ionisation has been instrumental in the study of proteins and large protein complexes[192-194].

Fourier transform ion cyclotron mass spectrometer

The principles of ion cyclotron mass spectrometry (ICR-MS) lie with the fact that the trajectory of ions are curved in a magnetic field; if this field is intense and the velocity of the ions are low, the radius of trajectory is small[167]. A balance of the centripetal and centrifugal force experienced by an ion in such a magnetic field will result in a stable trajectory

described by Equation 1.5, where *z* is the charge, *v* is velocity, *m* is mass and B is magnetic field.

$$z\mathrm{B} = \frac{mv}{r} \qquad [1.5]$$

Ions move in a circular trajectory of $2\pi r$ with a frequency (*f*) described by Equation 1.6, and the angular velocity (ω) is given by Equation 1.7. Combining Equations 1.5-1.7 it is clear ω is related to the *m/z* of the ion and hence ions of the same *m/z* have the same ICR frequency.

$$f = \frac{v}{2\pi r} \qquad [1.6]$$

$$\omega = 2\pi f = \frac{v}{r} = \frac{z}{m}B \qquad [1.7]$$

Ions circulate in tight packets in a cyclotron of small radius and can then be excited at their resonant cyclotron frequency through a potential applied to parallel plates orthogonal to the magnetic field (Figure 1.7), creating a larger ion cyclotron radius. The larger radius causes packets of ions of the same *m/z* to pass close to the detector plate, creating a charge on the pair of electrodes which is then digitised. In order to detect ions of a range of *m/z* at once, a broadband excitation can be applied. FT ICR-MS instruments have very high resolution and high sensitivity and when coupled with fragmentation techniques have provided great insight in proteomic studies and into the post-translational modifications of proteins[195-198], furthermore, coupling with labelling techniques, such as hydrogen-deuterium exchange, has provided insight into conformationally dynamic proteins[199]. Another advantage of FT ICR-MS instruments is that ions can be isolated in the cell and subjected to electron capture dissociation (ECD), see section 1.5.3.4.

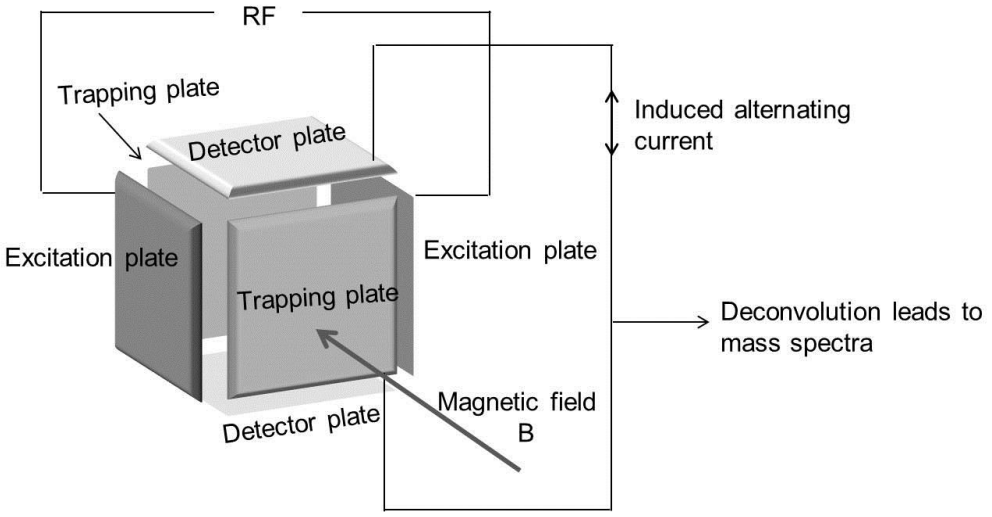

Figure 1.7: Schematic representation of the FT ICR-MS cell.

1.5.3.3 Detectors

Once the beam of ions has passed through the mass analyser it is has to be detected and transformed into a usable signal by the detector. There are two main types of detector which can be used within a mass spectrometer, those that involve direct measurement of the charges reaching the detector and those which involve increasing the intensity of the signal.

Microchannel plates

In TOF mass spectrometers microchannel plates (MCPs) are the most commonly used detector, MCPs are well suited to TOF instruments as they allow precise arrival times with narrow pulse widths to be obtained. A MCP is a plate comprising parallel channels coated in a semiconductor[200]. The ion input side of the detector is held at a negative potential of ~1 kV with respect to the output side. When an ion hits one of the semiconductor coated channels electrons are released, this process is continuous and results in a cascade of electrons which are accelerated by the potential gradient along the channel. Amplification of 10^2-10^4 is achieved for a single plate while connecting several plates together enable an amplification of 10^8 to be achieved[167]. The cascade of electrons are then measured as current.

FT ICR detector

As described above (section 1.5.2.2), a broad excitation frequency is applied to induce a larger cyclotron radius of ions of multiple *m/z*, which results in multiple signals being recorded on the detector plates at once. This signal is known as the free induction decay or transient and comprises a superposition of sine waves of different frequencies and amplitudes, depending on the ions, and is converted into useable spectra through a Fourier transformation.

1.5.3.4 Electron capture dissociation

Electron capture dissociation (ECD) is a non-ergodic method of fragmentation developed within the McLafferty laboratory[201], during which bond dissociation occurs much faster than typical bond vibration[202]. In ECD, low energy electrons are introduced to trapped gas-phase ions, producing a radical cation, $[M + nH]^{(n-1)+\bullet}$, which can then dissociate forming primarily c and z-type fragments[203]; however, b-type fragment ions have also been reported[204] (Figure 1.8). ECD is used with FT ICR-MS instrumentation and offers the advantage of inducing more general backbone fragmentation than other methods. ECD is an attractive fragmentation technique which can preserve post-translation modifications[167] and has provided insight in studies of protein unfolding[205,206].

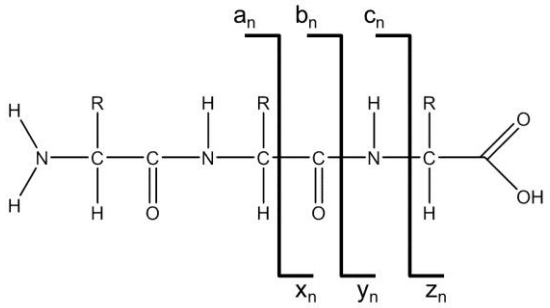

Figure 1.8: Protein fragment nomenclature.

1.5.4 Ion mobility-mass spectrometry

Ion mobility-mass spectrometry (IM-MS) is an electrophoretic gas-phase technique, which allows ions to be separated based not only on their mass (*m*) and charge (*z*) as in conventional mass spectrometry but also on their mobility (K) in a given buffer gas, which is related to their size and shape. There are three main types of instrumentation which can be coupled with mass spectrometry; linear drift tube (DT), travelling wave ion guide (TWIG) and field asymmetric ion mobility (FAIMS).

1.5.4.1 Drift tube ion mobility-mass spectrometry

Drift tube ion mobility spectrometry (DT IMS) is the traditional and simplest ion mobility instrumental configuration, the foundations of which were laid at the beginning of the 20th century in experimental[207,208] and theoretical[209] investigations of the movement of ions in gases, following which McDaniel *et al*[210,211] developed a low field drift tube in which gaseous ions drifted the length of the tube under the action of a weak field. These investigations formed the basis of DT IMS, which involves the measurement of an ion's velocity through a drift region attained under an electric field gradient whilst experiencing collisions with a buffer gas. From this measurement it is possible to determine the rotationally averaged collision cross section (CCS) of the species, synonymous with its size and shape. An ion of larger CCS will undergo a greater number of collisions with the buffer gas (usually helium) and hence will travel more slowly, whilst an ion of smaller CCS will undergo fewer collisions and reach the end of the drift region faster.

The behaviour of an ion travelling through a cell filled with buffer gas under the influence of an electric field (E) is dependent on ratio of the field strength to buffer gas number (N). At high values of E/N ions align in the field and their motion becomes dependent on E. At low values of E/N (low field) the motion of ions is more simplistic and ions possess low constant velocities proportional to E[212], as shown in Equation 1.8, where v_d is the drift velocity.

$$v_d = KE \quad [1.8]$$

The mobility of an ion (K) depends on its mass, charge (z) and shape, also known as the rotationally averaged collision cross section (Ω), and can be described by Equation 1.9, where μ is the reduced mass of the analyte and buffer gas, k_B is the Boltzman constant, T is the temperature reported in Kelvin and N is the density of the buffer gas.

$$K = \frac{3ze}{16N}\left(\frac{2\pi}{\mu k_B T}\right)^{\frac{1}{2}}\frac{1}{\Omega} \quad [1.9]$$

The mobility of an ion depends upon the temperature and pressure at which the measurements were taken and hence it is often reported as the reduced mobility (K_0), which accounts for temperature and pressure, as shown in Equation 1.10, enabling comparisons to be made at ease.

$$K_0 = K\left(\frac{T^0}{T}\right)\left(\frac{P}{P^0}\right) \quad [1.10]$$

Where, P^0 and T^0 are standard pressure (760 torr) and temperature (273.15 K) respectively.

In the early 1960s, the first reports of coupling ion mobility spectrometry and mass spectrometry creating the hybrid technique of DT IM-MS were reported[213-215]. DT IMS has subsequently been coupled with many different types of mass spectrometer, with the drift cell situated either before or after the mass analyser, the types of instrumentation and their subsequent applications have been reviewed elsewhere[216,217]. DT IM-MS has provided insightful results in the study of protein fold and unfolding[218-221], conformational implications of binding[222-224] and in aggregating systems[158,225,226]. The use of DT IM-MS instruments has traditionally been confined to research and development labs in which they could be built, however, in 2013 Agilent Technologies (Santa Clara, USA) introduced the first commercially available DT IM-MS which will surely prompt an increase in the use and applications of DT IM-MS.

1.5.4.2 Travelling wave ion guide mobility mass spectrometry

Waters (Manchester, UK) introduced the first commercially available integrated IM-MS instrument in 2006, the Synapt HDMS (high definition mass spectrometer) and have since introduced two further models. Synapt instruments use a technique known as travelling wave ion guide (TWIG) ion mobility, incorporating a travelling voltage wave into a RF guide[227,228]. In Synapt instruments, ions are first stored in a region known as the ion guide before being injected into the mobility cell, after separation in the mobility cell they travel

through an additional ion guide which transfers ions into the TOF. The mobility cell is filled with nitrogen buffer gas and comprises a series of planar electrodes arranged orthogonally to the ion beam, ions are radially confined within the cell due to an RF field applied to consecutive electrodes. Ions traverse the length of the cell due to a travelling wave, comprising a series of transient DC voltages which are superimposed on top of the RF voltage. Ions with the highest mobility travel on the crest of the wave whilst lower mobility ions will fall behind successive wave fronts before finally reaching the end of the ion guide. Unlike DT IM-MS, the time taken for an ion to traverse the mobility cell is not directly related to the CCS of that ion; however, is proportional to t_D^X, where X is an experimentally determined parameter dependent on instrumental settings. In order to obtain CCS using TWIG IM-MS one has to first measure the drift time for a calibration set of standards for comparison to drift times recorded previously on a DT IM-MS instrument, after which the CCS of the system under investigation can be calculated[229,230]. The commercial availability of TWIG IM-MS has provided access to a wide range of research groups and hence the applications of TWIG IM-MS to biological studies is varied, including the study of dynamic protein conformations[231] and protein:protein complexes[153,155,232] as well as in imaging[233], to name only a few. A recent report has, however, shown that for conformational or structural studies careful control over instrumental parameters must be taken to avoid thermal unfolding of ions[234].

1.5.4.3 Field asymmetric ion mobility-mass spectrometry

The final type of ion mobility instrumentation to be successfully coupled with MS is field asymmetric ion mobility (FAIMS). FAIMS was first reported in the late 1990s by Purves *et al*[235,236] and here ions are passed between a pair of electrodes under the influence of an opposing gas flow. An asymmetric waveform is applied to the upper electrode, comprising a high-voltage component and a longer low-voltage component of opposite polarity, whilst the lower electrode is held at ground potential. In order for ions to pass through the electrodes without collision, a DC 'compensation voltage' which reflects the difference in the high and low field mobility of the ion is applied, this compensation voltage can be thought of as a separation parameter. In FAIMS the velocity with which ions traverse the mobility region is not directly proportional to applied field and is instead dependent upon it. CCS cannot be directly determined from this IMS technique. Separation in FAIMS occurs at room temperature and at atmospheric pressure and hence is often coupled to the front end of a mass spectrometer. FAIMS devices are commercially available from Thermo Scientific (San Jose, USA), Ab Sciex (Concord, Canada) and Agilent technologies (Santa Clara, USA) in

collaboration with Owlstone (Cambridge, UK). The heating of ions in the high field of the FAIMS device makes it unsuitable for conformational studies; however, it has found use in the separation of complex mixtures and post-translationally modified peptides[237,238].

1.5.4.4 Estimation of theoretical collision cross sections

One particularly valuable feature of IM-MS for conformational studies is that it is often possible to determine a theoretical CCS to compare with experimental CCS. For protein and peptide studies this is most frequently achieved using structures obtained from NMR, X-ray crystallography or molecular modelling. Structures solved by NMR and X-ray crystallography are available online from the Protein Data Bank (PDB) which is run by the Research Collaboratory for Structural Bioinformatics[239]. Most frequently, coordinates either from PDB files or molecular modelling are input into a Fortran program developed by Shvartsburg and Jarrold known as MOBCAL (a program to CALculate MOBilities)[240-242], which can determine the rotationally averaged collision cross section for molecule. There are four different models for the ion-buffer gas collisions; projection approximation (PA), exact hard sphere scattering (EHSS), trajectory method (TM) and the projection superposition approximation (PSA) method.

In PA, the rotationally averaged CCS is calculated considering each possible orientation of the molecule[243]. The ion is modelled using a collection of overlapping hard spheres, whose radii are equal to the hard sphere collision distances. Buffer gas collisions are also considered purely as hard-sphere collisions and PA does not consider multiple collision events or long range interactions, which simplifies this model greatly; however, CCS determined from PA are often significantly underestimated[244]. The EHSS is similar to the PA model; however, it accounts for multiple collision processes as well as scattering[240]. The trajectory method was further developed to consider the ion as a collection of atoms and accounts for long-range interactions and close collisions between the ion and the buffer gas atoms as well as multiple collision events and is considered the most reliable method to determine CCS for large ions[241]. All theoretical CCS presented in this thesis were determined using the trajectory method. The TM method is computationally expensive and can therefore be slow, especially for large proteins and protein complexes, which prompted the development of the projection superposition approximation (PSA) model. In PSA CCS are determined in a similar fashion to that used in PA; however, collective shape and size effects are accounted for[245-248]. PSA is therefore more accurate than PA, in addition to being much faster than the TM.

1.5.5 Circular dichroism spectroscopy

Circular dichroism spectroscopy (CD) is a low resolution solution-phase spectroscopic technique, which can be used to estimate the secondary structural contents of peptides and proteins[249,250]. CD is based upon the differential absorption of left and right circularly polarised light by chiral chromophores. Proteins and peptides can contain a number of chiral chromophores, indeed in the far ultra violet (UV) region (280-180 nm) CD shows a number of characteristic features, including peptide bond absorption with a weak, broad $n \rightarrow \pi^*$ transition (centred around 210 nm) and a more intense $\pi \rightarrow \pi^*$ transition (centred around 190 nm). β-sheets, α-helices, turns and disordered regions all present distinctive spectra and hence the secondary structural content can be determined, using algorithms which compare the CD spectra obtained to structures solved by X-ray crystallography[250,251].

1.5.6 Transmission electron microscopy

Transmission electron microscopy (TEM) is a method of microscopy in which a beam of electrons is transmitted through a thin sample. TEM has the advantage of much higher resolving power (defined as half the wavelength of the illuminating beam) than traditional microscopic techniques due to the small wavelength of the electron beam[252]. In TEM a high voltage is applied to a tungsten filament (the cathode) creating electrons, which are accelerated towards a metal plate with a central aperture (the anode), some electrons pass through the aperture creating the electron beam. The electron beam is subsequently focused by a magnetic lens before passing through the sample. The majority of electrons will pass through the specimen without deviation, however, some will be scattered by heavy atoms within the sample[252]. The scattering causes a pattern to be formed in the emerging beam, which is refocused and magnified by post-sample lenses and is then viewed on the detector, typically a fluorescent screen or charge-coupled camera.

The higher the atomic number of the atoms within the sample, the more electrons are scattered and provide a higher contrast in the TEM images. As biological samples are generally composed of low atomic number atoms, heavy metal salts are often used as stains to enhance the contrast[253]. A commonly used stain for biological samples is uranyl acetate, which has been shown to cross-link and stabilise fragile macromolecular assemblies[253]. Uranyl acetate (UA) is a negative stain, meaning the stain dries around the molecules resulting in the area around the species of interest appearing dark whilst molecules themselves appear light. UA, however, can chemically react with biomolecules causing positive staining, where molecules appears dark on a clear background. TEM has proved

incredibly useful in the study of aggregating proteins[254], particularly those which form amyloid fibrils[255], afforded through the application of protein solutions to coated carbon grids which can be inserted into the electron microscope.

1.5.7 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a biophysical technique used to characterise the thermodynamics of binding interactions[256]. An isothermal titration calorimeter is most commonly based on a feedback network, which constantly measures the temperature change between a sample cell and a reference cell. A constant power is applied to the reference cell to activate the feedback loop which applies a variable power to the sample cell, in order to keep the temperature difference between the two cells small. If a reaction occurs within the sample cell, the temperature in the cell will change resulting in a change in the feedback loop to compensate for this, in an exothermic reaction the feedback power will decrease and in an endothermic reaction the feedback power will increase. Measurement of the deviation from baseline allows reaction characteristics to be determined. Strict controls are required to discount heat of change due to protein and ligand dilution and heat of mixing.

In a typical ITC experiment to study ligand binding to a peptide or protein, a cell is filled with the peptide/protein solution and a syringe is filled with the ligand at a high concentration. The ligand is then titrated in to the cell and the calorimeter measures the heat flow resulting from the heat effect induced by the change in composition at constant temperature. ITC experiments can be used to determine the association constant (K_a), stoichiometry (n), free energy (ΔG^0), enthalpy (ΔH^0) and entropy (ΔS^0) of binding[256], the analysis and equations through which these can be determined can be found in Appendix 1.2.

1.6 Summary

MS and IM-MS based methods have great promise for the study of multiple dynamic conformations of proteins and peptides, overcoming some of the major challenges of traditional techniques. This thesis investigates the ability of MS and DT IM-MS in combination with other biophysical methods such as CD, TEM and ITC to study the structure and interactions of peptides and proteins. Firstly, conformationally dynamic systems are investigated considering a peptide system which undergoes a disorder-order transition upon binding its partner (c-MYC:MAX) and a metamorphic protein (lymphotactin), (Chapters 3 and 4, respectively). The interactions occurring between a model glycosaminoglycan and lymphotactin, in addition to specific mutants are then studied

(Chapter 5). Finally, the interactions occurring between the model glycosaminoglycan and two members of the human β-defensin sub class of antimicrobial peptides are considered (Chapter 6). The ability of this combination of techniques to report on structure and interactions, and their future applicability to other dynamic systems are discussed.

1.7 References

(1) Buxbaum, E. *Fundamentals of protein structure and function*; Springer, 2007.
(2) Pauling, L. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*; Cornell University Press, 1960; Vol. 18.
(3) Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*; Macmillan, 1999.
(4) Richardson, J. S., *Adv. Protein Chem*, **1981**, *34*, 339.
(5) Bragg, L.; Kendrew, J. C.; Perutz, M. F., *Proc. R. Soc. Lond., A*, **1950**, *203*, 321-357.
(6) Pauling, L.; Corey, R. B.; Branson, H. R., *Proc. Natl. Acad. Sci. U. S. A.*, **1951**, *37*, 205-211.
(7) Pace, N. C.; Scholtz, M. J., *Biophys. J.*, **1998**, *75*, 422-427.
(8) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z., *J. Mol. Graph. Model.*, **2001**, *19*, 26-59.
(9) Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K., *Biophys. J.*, **2007**, *92*, 1439-1456.
(10) Williams, R. M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E. C.; Young, J.; Takayama, S.; Brown, C. J.; Dunker, A. K., *Pac. Symp. Biocomput.*, **2001**, 89-100.
(11) Fischer, E., *Ber. Dtsch. Chem. Ges.*, **1894**, *27*, 2985-2993.
(12) Wright, P. E.; Dyson, H. J., *J. Mol. Biol.*, **1999**, *293*, 321-331.
(13) Dyson, H. J.; Wright, P. E., *Nat. Rev. Mol. Cell. Biol.*, **2005**, *6*, 197-208.
(14) Murzin, A. G., *Science*, **2008**, *320*, 1725-1726.
(15) Goodchild, S.; Curmi, P. G.; Brown, L., *Biophys Rev*, **2011**, *3*, 143-153.
(16) Ha, J.-H.; Loh, S. N., *Chem. Euro. J.*, **2012**, *18*, 7984-7999.
(17) Vita, M.; Henriksson, M. In *Semin Cancer Biol*; Elsevier: 2006; Vol. 16, p 318-330.
(18) Blackwood, E. M.; Eisenman, R. N., *Science*, **1991**, *251*, 1211-1217.
(19) Lavigne, P.; Crump, M. P.; Gagne, S. M.; Hodges, R. S.; Kay, C. M.; Sykes, B. D., *J. Mol. Biol.*, **1998**, *281*, 165-181.
(20) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L., *Science*, **1988**, *240*, 1759-1764.
(21) Lavigne, P.; Kondejewski, L. H.; Houston Jr, M. E.; Sönnichsen, F. D.; Lix, B.; Sykes, B. D.; Hodges, R. S.; Kay, C. M., *J. Mol. Biol.*, **1995**, *254*, 505-520.
(22) Muhle-Goll, C.; Gibson, T.; Schuck, P.; Schubert, D.; Nalis, D.; Nilges, M.; Pastore, A., *Biochemistry*, **1994**, *33*, 11296-11306.
(23) Nair, S. K.; Burley, S. K., *Cell*, **2003**, *112*, 193-205.
(24) Pabo, C. O.; Sauer, R. T., *Annu. Rev. Biochem.*, **1992**, *61*, 1053-1095.
(25) Adhikary, S.; Eilers, M., *Nat. Rev. Mol. Cell. Biol.*, **2005**, *6*, 635 - 645.
(26) Blackwood, E. M.; Lüscher, B.; Eisenman, R. N., *Genes Dev.*, **1992**, *6*, 71-80.
(27) Li, Z.; Van Calcar, S.; Qu, C.; Cavenee, W. K.; Zhang, M. Q.; Ren, B., *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, *100*, 8164-8169.
(28) Fernandez, P. C.; Frank, S. R.; Wang, L.; Schroeder, M.; Liu, S.; Greene, J.; Cocito, A.; Amati, B., *Genes Dev.*, **2003**, *17*, 1115-1129.
(29) Cawley, S.; Bekiranov, S.; Ng, H. H.; Kapranov, P.; Sekinger, E. A.; Kampa, D.; Piccolboni, A.; Sementchenko, V.; Cheng, J.; Williams, A. J., *Cell*, **2004**, *116*, 499-509.
(30) Gearhart, J.; Pashos, E. E.; Prasad, M. K., *N. Engl. J. Med.*, **2007**, *357*, 1469-1472.
(31) Eisenman, R. N. *The Myc/Max/Mad Transcription Factor Network*; Springer, 2006; Vol. 302.
(32) Meyer, N.; Penn, L. Z., *Nat. Rev. Cancer*, **2008**, *8*, 976-990.
(33) Field, J. K.; Spandidos, D., *Anticancer Res.*, **1990**, *10*, 1.
(34) Li, Q.; Dang, C. V., *Mol. Cell. Biol.*, **1999**, *19*, 5339-5351.

(35) Nesbit, C. E.; Tersak, J. M.; Prochownik, E. V., *Oncogene*, **1999**, *18*, 3004-3016.
(36) Ponzielli, R.; Katz, S.; Barsyte-Lovejoy, D.; Penn, L. Z., *Eur. J. Cancer*, **2005**, *41*, 2485-2501.
(37) Hermeking, H., *Curr. Cancer Drug Targets*, **2003**, *3*, 163-175.
(38) Sodir, N.; Evan, G., *J. Biol.*, **2009**, *8*, 77.
(39) Soucek, L.; Whitfield, J.; Martins, C.; Finch, A.; Murphy, D.; Sodir, N.; Karnezis, A.; Swigart, L.; Nasi, S.; Evan, G., *Nature*, **2008**, *455*, 679 - 683.
(40) Hammoudeh, D. I.; Follis, A. V.; Prochownik, E. V.; Metallo, S. J., *J. Am. Chem. Soc.*, **2009**, *131*, 7390-7401.
(41) Wang, H.; Hammoudeh, D. I.; Follis, A. V.; Reese, B. E.; Lazo, J. S.; Metallo, S. J.; Prochownik, E. V., *Mol. Cancer Ther.*, **2007**, *6*, 2399-2408.
(42) Wang, H.; Chauhan, J.; Hu, A.; Pendleton, K.; Yap, J. L.; Sabato, P. E.; Jones, J. W.; Perri, M.; Yu, J.; Cione, E., *Oncotarget*, **2013**, *4*, 936.
(43) Baggiolini, M.; Dewald, B.; Moser, B., *Annu. Rev. Immunol.*, **1997**, *15*, 675-705.
(44) Dammann, O.; OShea, T., *J. Interferon Chytokine Res.*, **2002**, *22*, 1067-1068.
(45) Zlotnik, A.; Yoshie, O., *Immunity*, **2000**, *12*, 121-127.
(46) Kim, K.-S.; Rajarathnam, K.; Clark-Lewis, I.; Sykes, B. D., *FEBS Lett.*, **1996**, *395*, 277-282.
(47) Mizoue, L. S.; Bazan, J. F.; Johnson, E. C.; Handel, T. M., *Biochemistry*, **1999**, *38*, 1402-1414.
(48) Lodi, P. J.; Garrett, D. S.; Kuszewski, J.; Tsang, M.-w.; Weatherbee, J. A.; Leonard, W. J.; Gronenborn, A. M.; Clore, G., *Science*, **1994**, *263*, 1762-1767.
(49) Fernandez, E. J.; Lolis, E., *Annu. Rev. Pharmacol. Toxicol.*, **2002**, *42*, 469-499.
(50) Allen, S. J.; Crown, S. E.; Handel, T. M., *Annu. Rev. Immunol.*, **2007**, *25*, 787-820.
(51) Rollins, B. J., *Blood*, **1997**, *90*, 909-928.
(52) Laing, K. J.; Secombes, C. J., *Dev. Comp. Immunol.*, **2004**, *28*, 443-460.
(53) Handel, T.; Hamel, D. *Chemokines, Part A*; Access Online via Elsevier, 2009.
(54) Rot, A.; von Andrian, U. H., *Annu. Rev. Immunol.*, **2004**, *22*, 891-928.
(55) Luster, A. D., *Curr. Opin. Imunol.*, **2002**, *14*, 129-135.
(56) Rossi, D.; Zlotnik, A., *Annu. Rev. Immunol.*, **2000**, *18*, 217-242.
(57) Baggiolini, M.; Loetscher, P., *Immunol. Today*, **2000**, *21*, 418-420.
(58) Baggiolini, M., *Nature*, **1998**, *392*, 565-568.
(59) Proudfoot, A. E.; Buser, R.; Borlat, F.; Alouani, S.; Soler, D.; Offord, R. E.; Schröder, J.-M.; Power, C. A.; Wells, T. N., *J. Biol. Chem.*, **1999**, *274*, 32478-32485.
(60) Moser, B., *Eur. Cytokine Netw.*, **2003**, *14*, 204-210.
(61) Hoogewerf, A. J.; Kuschert, G. S.; Proudfoot, A. E.; Borlat, F.; Clark-Lewis, I.; Power, C. A.; Wells, T. N., *Biochemistry*, **1997**, *36*, 13570-13578.
(62) Proudfoot, A. E. I.; Handel, T. M.; Johnson, Z.; Lau, E. K.; LiWang, P.; Clark-Lewis, I.; Borlat, F.; Wells, T. N. C.; Kosco-Vilbois, M. H., *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, *100*, 1885-1890.
(63) Handel, T. M.; Johnson, Z.; Crown, S. E.; Lau, E. K.; Proudfoot, A. E., *Annu. Rev. Biochem.*, **2005**, *74*, 385-410.
(64) Murphy, P. M.; Baggiolini, M.; Charo, I. F.; Hebert, C. A.; Horuk, R.; Matsushima, K.; Miller, L. H.; Oppenheim, J. J.; Power, C. A., *Pharmacol. Rev.*, **2000**, *52*, 145-176.
(65) Charo, I. F.; Taubman, M. B., *Circ. Res.*, **2004**, *95*, 858-866.
(66) Gerard, C.; Rollins, B. J., *Nat. Immunol.*, **2001**, *2*, 108-115.
(67) Kelner, G.; Kennedy, J.; Bacon, K.; Kleyensteuber, S.; Largaespada, D.; Jenkins, N.; Copeland, N.; Bazan, J.; Moore, K.; Schall, T., *Science*, **1994**, *266*, 1395-1399.
(68) Volkman, B. F.; Liu, T. Y.; Peterson, F. C., *Methods Enzymol.*, **2009**, *461*, 51-70.
(69) Kuloglu, E. S.; McCaslin, D. R.; Kitabwalla, M.; Pauza, C. D.; Markley, J. L.; Volkman, B. F., *Biochemistry*, **2001**, *40*, 12486-12496.
(70) Kuloglu, E. S.; McCaslin, D. R.; Markley, J. L.; Volkman, B. F., *J. Biol. Chem.*, **2002**, *277*, 17863-17870.
(71) Tuinstra, R. L.; Peterson, F. C.; Kutlesa, S.; Elgin, E. S.; Kron, M. A.; Volkman, B. F., *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105*, 5057-5062.
(72) Yoshida, T.; Imai, T.; Kakizaki, M.; Nishimura, M.; Yoshie, O., *FEBS Lett.*, **1995**, *360*, 155-159.
(73) Kennedy, J.; Kelner, G. S.; Kleyensteuber, S.; Schall, T. J.; Weiss, M. C.; Yssel, H.; Schneider, P. V.; Cocks, B. G.; Bacon, K. B.; Zlotnik, A., *J. Immunol.*, **1995**, *155*, 203-209.

(74) Tuinstra, R. L.; Peterson, F. C.; Elgin, E. S.; Pelzek, A. J.; Volkman, B. F., *Biochemistry*, **2007**, *46*, 2564-2573.
(75) Tyler, R. C.; Murray, N. J.; Peterson, F. C.; Volkman, B. F., *Biochemistry*, **2011**, *50*, 7077-7079.
(76) Hedrick, J. A.; Zlotnik, A. In *Methods Enzymol.*; Richard, H., Ed.; Academic Press: 1997; Vol. Volume 287, p 206-215.
(77) Müller, S.; Dorner, B.; Korthäuer, U.; Mages, H. W.; D'Apuzzo, M.; Senger, G.; Kroczek, R. A., *Eur. J. Immunol.*, **1995**, *25*, 1744-1748.
(78) Boismenu, R.; Feng, L.; Xia, Y. Y.; Chang, J.; Havran, W. L., *J. Immunol.*, **1996**, *157*, 985-992.
(79) Tikhonov, I.; Kitabwalla, M.; Wallace, M.; Malkovsky, M.; Volkman, B.; Pauza, C. D., *Cytokine*, **2001**, *16*, 73-78.
(80) Yoshida, T.; Imai, T.; Takagi, S.; Nishimura, M.; Ishikawa, I.; Yaoi, T.; Yoshie, O., *FEBS Lett.*, **1996**, *395*, 82-88.
(81) Dilloo, D.; Bacon, K.; Holden, W.; Zhong, W.; Burdach, S.; Zlotnik, A.; Brenner, M., *Nat. Med.*, **1996**, *2*, 1090-1095.
(82) Huang, H.; Bi, X.; Yuan, J.; Xu, S.; Guo, X.; Xiang, J., *Gene Ther.*, **2005**, *12*, 999-1010.
(83) Rousseau, R. F.; Haight, A. E.; Hirschmann-Jax, C.; Yvon, E. S.; Rill, D. R.; Mei, Z.; Smith, S. C.; Inman, S.; Cooper, K.; Alcoser, P., *Blood*, **2003**, *101*, 1718-1726.
(84) Taylor, K.; Barran, P. E.; Dorin, J. R., *Biopolymers*, **2008**, *90*, 1-7.
(85) Sahl, H.-G.; Pag, U.; Bonness, S.; Wagner, S.; Antcheva, N.; Tossi, A., *J. Leukoc. Biol.*, **2005**, *77*, 466-475.
(86) Hoffmann, J. A.; Hetru, C., *Immunol. Today*, **1992**, *13*, 411-415.
(87) Broekaert, W. F.; Terras, F.; Cammue, B.; Osborn, R. W., *Plant Physiol.*, **1995**, *108*, 1353.
(88) Raj, P. A.; Dentino, A. R., *FEMS Microbiol. Lett.*, **2002**, *206*, 9-18.
(89) Tang, Y.-Q.; Yuan, J.; Ösapay, G.; Ösapay, K.; Tran, D.; Miller, C. J.; Ouellette, A. J.; Selsted, M. E., *Science*, **1999**, *286*, 498-502.
(90) Selsted, M. E., *Curr. Protein. Pept. Sci.*, **2004**, *5*, 365-371.
(91) Linzmeier, R. M.; Ganz, T., *Genomics*, **2005**, *86*, 423-430.
(92) Schutte, B. C.; Mitros, J. P.; Bartlett, J. A.; Walters, J. D.; Jia, H. P.; Welsh, M. J.; Casavant, T. L.; McCray, P. B., Jr., *Proc. Natl. Acad. Sci. U. S. A.*, **2002**, *99*, 2129-2133.
(93) Hoover, D. M.; Rajashankar, K. R.; Blumenthal, R.; Puri, A.; Oppenheim, J. J.; Chertov, O.; Lubkowski, J., *J. Biol. Chem.*, **2000**, *275*, 32911-32918.
(94) Sawai, M. V.; Jia, H. P.; Liu, L.; Aseyev, V.; Wiencek, J. M.; McCray, P. B.; Ganz, T.; Kearney, W. R.; Tack, B. F., *Biochemistry*, **2001**, *40*, 3810-3816.
(95) Hoover, D. M.; Chertov, O.; Lubkowski, J., *J. Biol. Chem.*, **2001**, *276*, 39021-39026.
(96) Schibli, D. J.; Hunter, H. N.; Aseyev, V.; Starner, T. D.; Wiencek, J. M.; McCray, P. B., Jr.; Tack, B. F.; Vogel, H. J., *J. Biol. Chem.*, **2002**, *277*, 8279-8289.
(97) Bauer, F.; Schweimer, K.; Klüver, E.; Conejo-Garcia, J.-R.; Forssmann, W.-G.; Rösch, P.; Adermann, K.; Sticht, H., *Protein Sci.*, **2001**, *10*, 2470-2479.
(98) Landon, C.; Thouzeau, C.; Labbé, H.; Bulet, P.; Vovelle, F., *J. Biol. Chem.*, **2004**, *279*, 30433-30439.
(99) Selsted, M. E.; Ouellette, A. J., *Nat. Immunol.*, **2005**, *6*, 551-557.
(100) Bensch, K. W.; Raida, M.; Mägert, H.-J.; Schulz-Knappe, P.; Forssmann, W.-G., *FEBS Lett.*, **1995**, *368*, 331-335.
(101) Zhao, C.; Wang, I.; Lehrer, R. I., *FEBS Lett.*, **1996**, *396*, 319-322.
(102) Harder, J.; Bartels, J.; Christophers, E.; Schroder, J., *Nature*, **1997**, *387*, 861-861.
(103) Harder, J.; Bartels, J.; Christophers, E.; Schröder, J.-M., *J. Biol. Chem.*, **2001**, *276*, 5707-5713.
(104) Devine, D. A.; Hancock, R. E. W. *Mammalian Host Defense Peptides*; Cambridge university press: Cambridge, 2004.
(105) Lehrer, R. I., *Nat. Rev. Microbiol.*, **2004**, *2*, 727-738.
(106) Ganz, T., *Science*, **1999**, *286*, 420-421.
(107) Weinberg, A.; Krisanaprakornkit, S.; Dale, B., *Crit. Rev. Oral Biol. Med.*, **1998**, *9*, 399-414.
(108) Schroeder, B. O.; Wu, Z.; Nuding, S.; Groscurth, S.; Marcinowski, M.; Beisner, J.; Buchner, J.; Schaller, M.; Stange, E. F.; Wehkamp, J., *Nature*, **2011**, *469*, 419-423.
(109) Lehrer, R. I., *Nature*, **2011**, *469*, 309-310.
(110) Wu, Z., *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, *100*, 8880-8885.

(111) Campopiano, D. J.; Clarke, D. J.; Polfer, N. C.; Barran, P. E.; Langley, R. J.; Govan, J. R. W.; Maxwell, A.; Dorin, J. R., *J. Biol. Chem.*, **2004**, *279*, 48671-48679.
(112) Brogden, K. A., *Nat. Rev. Microbiol.*, **2005**, *3*, 238-250.
(113) Goldman, M. J.; Anderson, G. M.; Stolzenberg, E. D.; Kari, U. P.; Zasloff, M.; Wilson, J. M., *Cell*, **1997**, *88*, 553-560.
(114) Bals, R.; Wang, X.; Wu, Z.; Freeman, T.; Bafna, V.; Zasloff, M.; Wilson, J. M., *J. Clin. Invest.*, **1998**, *102*, 874.
(115) Shafer, W. M. *Antimicrobial peptides and human disease*; Springer, 2006.
(116) Shai, Y., *Trends Biochem. Sci.*, **1995**, *20*, 460-464.
(117) Oren, Z.; Shai, Y., *Eur. J. Biochem.*, **1996**, *237*, 303-310.
(118) Oren, Z.; Shai, Y., *Pept. Sci.*, **1998**, *47*, 451-463.
(119) Sass, V.; Pag, U.; Tossi, A.; Bierbaum, G.; Sahl, H.-G., *Int. J. Med. Microbiol.*, **2008**, *298*, 619-633.
(120) Sass, V.; Schneider, T.; Wilmes, M.; Körner, C.; Tossi, A.; Novikova, N.; Shamova, O.; Sahl, H.-G., *Infect. Immun.*, **2010**, *78*, 2793-2800.
(121) Krishnakumari, V.; Nagaraj, R., *Biochim. Biophys. Acta. Biomembranes.*, **2012**, *1818*, 1386-1394.
(122) Yang, D.; Chertov, O.; Bykovskaia, S.; Chen, Q.; Buffo, M.; Shogan, J.; Anderson, M.; Schröder, J.; Wang, J.; Howard, O., *Science*, **1999**, *286*, 525-528.
(123) Jin, G.; Kawsar, H. I.; Hirsch, S. A.; Zeng, C.; Jia, X.; Feng, Z.; Ghosh, S. K.; Zheng, Q. Y.; Zhou, A.; McIntyre, T. M., *PLoS ONE*, **2010**, *5*, e10993.
(124) Röhrl, J.; Yang, D.; Oppenheim, J. J.; Hehlgans, T., *J. Immunol.*, **2010**, *184*, 6688-6694.
(125) Seo, E. S.; Blaum, B. S.; Vargues, T.; De Cecco, M.; Deakin, J. A.; Lyon, M.; Barran, P. E.; Campopiano, D. J.; Uhrin, D., *Biochemistry*, **2010**, *49*, 10486-10495.
(126) Taylor, K.; Clarke, D. J.; McCullough, B.; Chin, W.; Seo, E.; Yang, D.; Oppenheim, J.; Uhrin, D.; Govan, J. R. W.; Campopiano, D. J.; MacMillan, D.; Barran, P.; Dorin, J. R., *J. Biol. Chem.*, **2008**, *283*, 6631-6639.
(127) Tyrrell, C.; De Cecco, M.; Reynolds, N. L.; Kilanowski, F.; Campopiano, D.; Barran, P.; Macmillan, D.; Dorin, J. R., *Mol. Immunol.*, **2010**, *47*, 1378-1382.
(128) Weinberg, A.; Jin, G.; Sieg, S.; McCormick, T. S., *Front. Immunol.*, **2012**, *3*.
(129) Kawsar, H. I.; Weinberg, A.; Hirsch, S. A.; Venizelos, A.; Howell, S.; Jiang, B.; Jin, G., *Oral Oncol.*, **2009**, *45*, 696-702.
(130) Schröder, J.-M.; Harder, J., *Int. J. Biochem. Cell Biol.*, **1999**, *31*, 645-651.
(131) Sørensen, O. E.; Thapa, D. R.; Roupé, K. M.; Valore, E. V.; Sjöbring, U.; Roberts, A. A.; Schmidtchen, A.; Ganz, T., *J. Clin. Invest.*, **2006**, *116*, 1878-1885.
(132) Nomura, I.; Goleva, E.; Howell, M. D.; Hamid, Q. A.; Ong, P. Y.; Hall, C. F.; Darst, M. A.; Gao, B.; Boguniewicz, M.; Travers, J. B., *J. Immunol.*, **2003**, *171*, 3262-3269.
(133) Dyson, H. J.; Wright, P. E., *Chem. Rev.*, **2004**, *104*, 3607-3622.
(134) Ishima, R.; Torchia, D. A., *Nat. Struct. Mol. Biol.*, **2000**, *7*, 740-743.
(135) Clore, G. M.; Gronenborn, A. M., *Trends Biotechnol.*, **1998**, *16*, 22-34.
(136) Mesters, J. *Principles of protein X-ray crystallography*; Springer, 2007.
(137) Usón, I.; Sheldrick, G. M., *Curr. Opin. Struct. Biol.*, **1999**, *9*, 643-648.
(138) Chowdhury, S. K.; Katta, V.; Chait, B. T., *J. Am. Chem. Soc.*, **1990**, *112*, 9012-9013.
(139) Konermann, L.; Douglas, D. J., *Biochemistry*, **1997**, *36*, 12296-12302.
(140) Dobo, A.; Kaltashov, I. A., *Anal. Chem.*, **2001**, *73*, 4763-4773.
(141) Kuprowski, M. C.; Konermann, L., *Anal. Chem.*, **2007**, *79*, 2499-2506.
(142) Boyarkin, O. V.; Mercier, S. R.; Kamariotis, A.; Rizzo, T. R., *J. Am. Chem. Soc.*, **2006**, *128*, 2816-2817.
(143) Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R., *J. Am. Chem. Soc.*, **2007**, *129*, 13820-13821.
(144) Stearns, J. A.; Seaiby, C.; Boyarkin, O. V.; Rizzo, T. R., *Phys. Chem. Chem. Phys.*, **2009**, *11*, 125-132.
(145) Pagel, K.; Kupser, P.; Bierau, F.; Polfer, N. C.; Steill, J. D.; Oomens, J.; Meijer, G.; Koksch, B.; von Helden, G., *Int. J. Mass. Spectrom.*, **2009**, *283*, 161-168.
(146) Dugourd, P.; Antoine, R.; Breaux, G.; Broyer, M.; Jarrold, M. F., *J. Am. Chem. Soc.*, **2005**, *127*, 4675-4679.
(147) Jurneczko, E.; Barran, P. E., *Analyst*, **2011**.
(148) Breuker, K.; McLafferty, F. W., *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105*, 18145-18152.
(149) Loo, J. A., *Int. J. Mass. Spectrom.*, **2000**, *200*, 175-186.

(150) Pacholarz, K. J.; Garlish, R. A.; Taylor, R. J.; Barran, P. E., *Chem. Soc. Rev.*, **2012**, *41*, 4335-4355.
(151) Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S., *J. Am. Soc. Mass Spectrom.*, **2012**, *23*, 431-441.
(152) Hopper, J. T. S.; Oldham, N. J., *J. Am. Soc. Mass Spectrom.*, **2009**, *20*, 1851-1858.
(153) Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V., *Nat. Protoc.*, **2008**, *3*, 1139-1152.
(154) Barrera, N. P.; Di Bartolo, N.; Booth, P. J.; Robinson, C. V., *Science*, **2008**, *321*, 243-246.
(155) Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; van Duijn, E.; Heck, A. J., *Nat. Chem.*, **2010**, *3*, 126-132.
(156) Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T., *Nat. Chem.*, **2011**, *3*, 172-177.
(157) Bernstein, S. L.; Wyttenbach, T.; Baumketner, A.; Shea, J.-E.; Bitan, G.; Teplow, D. B.; Bowers, M. T., *J. Am. Chem. Soc.*, **2005**, *127*, 2075-2084.
(158) Cole, H. L.; Kalapothakis, J. M. D.; Bennett, G.; Barran, P. E.; MacPhee, C. E., *Angew. Chem.*, **2010**, *122*, 9638-9641.
(159) Ouyang, Z.; Takáts, Z.; Blake, T. A.; Gologan, B.; Guymon, A. J.; Wiseman, J. M.; Oliver, J. C.; Davisson, V. J.; Cooks, R. G., *Science*, **2003**, *301*, 1351-1354.
(160) Barran, P. E.; Polfer, N. C.; Campopiano, D. J.; Clarke, D. J.; Langridge-Smith, P. R. R.; Langley, R. J.; Govan, J. R. W.; Maxwell, A.; Dorin, J. R.; Millar, R. P.; Bowers, M. T., *Int. J. Mass. Spectrom.*, **2005**, *240*, 273-284.
(161) Thomson, J. J., *Proc. R. Soc. Lond., A*, **1913**, *89*, 1-20.
(162) Aston, F. W., *Philos. Mag.*, **1919**, *38*, 707-714.
(163) Squires, G., *J. Chem. Soc. Dalton Trans.*, **1998**, 3893-3900.
(164) Heck, A. J. R., *Nat Meth*, **2008**, *5*, 927-933.
(165) Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G., *Methods*, **2005**, *35*, 211-222.
(166) Dinh, G. G. a. T. V. *Handbook of Spectroscopy*; Wiley VCH, 2003; Vol. 1.
(167) Edmond de Hoffmann; Stroobant, V. *Mass Spectrometry: Principles and Applications*; 3 ed.; John Wiley and Sons: Chichester, 2007.
(168) Dass, C. *Fundamentals of contemporary mass spectrometry*; Wiley. com, 2007; Vol. 16.
(169) Karas, M.; Bachmann, D.; Hillenkamp, F., *Anal. Chem.*, **1985**, *57*, 2935-2939.
(170) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T., *Rapid Commun. Mass Spectrom.*, **1988**, *2*, 151-153.
(171) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B., *J. Chem. Phys.*, **1968**, *49*, 2240-2249.
(172) Clegg, G.; Dole, M., *Biopolymers*, **1971**, *10*, 821-826.
(173) Dole, M.; Cox Jr, H.; Gieniec, J., *Adv. Chem. Ser*, **1973**, *125*, 73-84.
(174) Yamashita, M.; Fenn, J. B., *J. Phys. Chem.*, **1984**, *88*, 4451-4459.
(175) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C., *Science*, **1989**, *246*, 64-71.
(176) Fenn, J. B., *Angew. Chem.*, **2003**, *42*, 3871-3894.
(177) Kebarle, P.; Tang, L., *Anal. Chem.*, **1993**, *65*, 972A-986A.
(178) Wu, X.; Oleschuk, R. D.; Cann, N. M., *Analyst*, **2012**, *137*, 4150-4161.
(179) Rayleigh, L., *Philosophical Magazine Series 5*, **1882**, *14*, 184-186.
(180) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S., *Anal. Chem.*, **2012**, *85*, 2-9.
(181) Nguyen, S.; Fenn, J. B., *Proc. Natl. Acad. Sci. U. S. A.*, **2007**, *104*, 1111-1117.
(182) Iribarne, J.; Thomson, B., *J. Chem. Phys.*, **1976**, *64*, 2287.
(183) Kebarle, P.; Verkerk, U. H., *Mass Spectrom. Rev.*, **2009**, *28*, 898-917.
(184) Iavarone, A. T.; Williams, E. R., *J. Am. Chem. Soc.*, **2003**, *125*, 2319-2327.
(185) Ahadi, E.; Konermann, L., *J. Phys. Chem. B*, **2012**, *116*, 104-112.
(186) Konermann, L.; Rodriguez, A. D.; Liu, J., *Anal. Chem.*, **2012**, *84*, 6798-6804.
(187) Wilm, M. S.; Mann, M., *Int. J. Mass. Spectrom. Ion. Proc.*, **1994**, *136*, 167-180.
(188) Wilm, M.; Mann, M., *Anal. Chem.*, **1996**, *68*, 1-8.
(189) Juraschek, R.; Dülcks, T.; Karas, M., *J. Am. Soc. Mass Spectrom.*, **1999**, *10*, 300-308.
(190) Mamyrin, B.; Karataev, V.; Shmikk, D.; Zagulin, V., *Zh Eksp Teor Fiz*, **1973**, *64*, 82-89.
(191) van den Heuvel, R. H.; van Duijn, E.; Mazon, H.; Synowsky, S. A.; Lorenzen, K.; Versluis, C.; Brouns, S. J.; Langridge, D.; van der Oost, J.; Hoyes, J., *Anal. Chem.*, **2006**, *78*, 7473-7483.
(192) Van Berkel, W. J. H.; Van Den Heuvel, R. H. H.; Versluis, C.; Heck, A. J. R., *Protein Sci.*, **2000**, *9*, 435-439.

(193) Benesch, J. L. P.; Ruotolo, B.; Simmons, D. A.; Robinson, C. V., *Chem. Rev.*, **2007**, *107*, 3544-3567.
(194) Chamot-Rooke, J.; Mikaty, G.; Malosse, C.; Soyer, M.; Dumont, A.; Gault, J.; Imhaus, A.-F.; Martin, P.; Trellet, M.; Clary, G.; Chafey, P.; Camoin, L.; Nilges, M.; Nassif, X.; Duménil, G., *Science*, **2011**, *331*, 778-782.
(195) Cooper, H. J.; Heath, J. K.; Jaffray, E.; Hay, R. T.; Lam, T. T.; Marshall, A. G., *Anal. Chem.*, **2004**, *76*, 6982-6988.
(196) Renfrow, M. B.; Cooper, H. J.; Tomana, M.; Kulhavy, R.; Hiki, Y.; Toma, K.; Emmett, M. R.; Mestecky, J.; Marshall, A. G.; Novak, J., *J. Biol. Chem.*, **2005**, *280*, 19136-19145.
(197) André, M.; Le Caer, J.-P.; Greco, C.; Planchon, S.; El Nemer, W.; Boucheix, C.; Rubinstein, E.; Chamot-Rooke, J.; Le Naour, F., *Proteomics*, **2006**, *6*, 1437-1449.
(198) Chamot-Rooke, J.; Rousseau, B.; Lanternier, F.; Mikaty, G.; Mairey, E.; Malosse, C.; Bouchoux, G.; Pelicic, V.; Camoin, L.; Nassif, X.; Duménil, G., *Proc. Natl. Acad. Sci. U. S. A.*, **2007**, *104*, 14783-14788.
(199) Bou-Assaf, G. M.; Chamoun, J. E.; Emmett, M. R.; Fajer, P. G.; Marshall, A. G., *Int. J. Mass. Spectrom.*, **2011**, *302*, 116-124.
(200) Ladislas Wiza, J., *Nucl. Instrum. Methods*, **1979**, *162*, 587-601.
(201) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., *J. Am. Chem. Soc.*, **1998**, *120*, 3265-3266.
(202) Breuker, K.; Oh, H. B.; Lin, C.; Carpenter, B. K.; McLafferty, F. W., *Proc. Natl. Acad. Sci. U. S. A.*, **2004**, *101*, 14011-14016.
(203) Roepstorff, P.; Fohlman, J., *Biomed. Mass Spectrom.*, **1984**, *11*, 601.
(204) Cooper, H. J., *J. Am. Soc. Mass Spectrom.*, **2005**, *16*, 1932-1940.
(205) Oh, H.; Breuker, K.; Sze, S. K.; Ge, Y.; Carpenter, B. K.; McLafferty, F. W., *Proc. Natl. Acad. Sci. U. S. A.*, **2002**, *99*, 15863-15868.
(206) Breuker, K.; Oh, H. B.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W., *J. Am. Chem. Soc.*, **2002**, *124*, 6407-6420.
(207) Erikson, H. A., *Physical Review*, **1927**, *30*, 339-348.
(208) Bradbury, N. E., *Physical Review*, **1932**, *40*, 508-523.
(209) Langevin, P., *Annales De Chimie Et De Physique*, **1903**, *28*, 433-530.
(210) Barnes, W. S.; Martin, D. W.; McDaniel, E. W., *Phys. Rev. Lett.*, **1961**, *6*, 110-111.
(211) McDaniel, E. W.; Martin, D. W.; Barnes, W. S., *Rev. Sci. Instrum.*, **1962**, *33*, 2-7.
(212) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; Wiley-VCH Verlag GmbH & Co. KGaA, 2005.
(213) K. B. McAfee, J.; Edelson, D., *Proc. Phys. Soc.*, **1963**, *81*, 382.
(214) Hogg, A. M.; Kebarle, P., *J. Chem. Phys.*, **1965**, *43*, 449-456.
(215) Kebarle, P.; Hogg, A. M., *J. Chem. Phys.*, **1965**, *42*, 668-674.
(216) Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H., *J. Mass. Spectrom.*, **2008**, *43*, 1-22.
(217) Harvey, S. R.; MacPhee, C. E.; Barran, P. E., *Methods*, **2011**, *54*, 454-461.
(218) Myung, S.; Badman, E. R.; Lee, Y. J.; Clemmer, D. E., *J. Phys. Chem. A*, **2002**, *106*, 9976-9982.
(219) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., *J. Am. Chem. Soc.*, **1995**, *117*, 10141-10142.
(220) Pierson, N. A.; Valentine, S. J.; Clemmer, D. E., *J. Phys. Chem. B*, **2010**, *114*, 7777-7783.
(221) Kalapothakis, J. M. D.; Berezovskaya, Y.; Zampronio, C. G.; Faull, P. A.; Barran, P. E.; Cooper, H. J., *Chem. Comm.*, **2014**.
(222) Faull, P. A.; Korkeila, K. E.; Kalapothakis, J. M.; Gray, A.; McCullough, B. J.; Barran, P. E., *Int. J. Mass. Spectrom.*, **2009**, *283*, 140-148.
(223) Glover, M. S.; Dilger, J. M.; Zhu, F.; Clemmer, D. E., *Int. J. Mass. Spectrom.*, **2013**, *354*, 318-325.
(224) Grabenauer, M.; Bernstein, S. L.; Lee, J. C.; Wyttenbach, T.; Dupuis, N. F.; Gray, H. B.; Winkler, J. R.; Bowers, M. T., *J. Phys. Chem. B*, **2008**, *112*, 11147-11154.
(225) Gessel, M. M.; Bernstein, S.; Kemper, M.; Teplow, D. B.; Bowers, M. T., *ACS Chem. Neurosci.*, **2012**.
(226) Gessel, M. M.; Wu, C.; Li, H.; Bitan, G.; Shea, J.-E.; Bowers, M. T., *Biochemistry*, **2011**, *51*, 108-117.
(227) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H., *Rapid Commun. Mass Spectrom.*, **2004**, *18*, 2401-2414.

(228) Pringle, S.; Giles, K.; Wildgoose, J.; Williams, J.; Slade, S.; Thalassinos, K.; Bateman, R.; Bowers, M.; Scrivens, J., *Int. J. Mass. Spectrom.*, **2007**, *261*, 1-12.
(229) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V., *Nat. Protoc.*, **2008**, *3*, 1139-1152.
(230) Leary, J. A.; Schenauer, M. R.; Stefanescu, R.; Andaya, A.; Ruotolo, B. T.; Robinson, C. V.; Thalassinos, K.; Scrivens, J. H.; Sokabe, M.; Hershey, J. W. B., *J. Am. Soc. Mass Spectrom.*, **2009**, *20*, 1699-1706.
(231) Jenner, M.; Ellis, J.; Huang, W. C.; Lloyd Raven, E.; Roberts, G. C.; Oldham, N. J., *Angew. Chem.*, **2011**, *50*, 8291-8294.
(232) Smith, D. P.; Radford, S. E.; Ashcroft, A. E., *Proc. Natl. Acad. Sci. U. S. A.*, **2010**, *107*, 6794-6798.
(233) Stauber, J.; MacAleese, L.; Franck, J.; Claude, E.; Snel, M.; Kaletas, B. K.; Wiel, I. M. V. D.; Wisztorski, M.; Fournier, I.; Heeren, R. M. A., *J. Am. Soc. Mass Spectrom.*, **2010**, *21*, 338-347.
(234) Merenbloom, S. I.; Flick, T. G.; Williams, E. R., *J. Am. Soc. Mass Spectrom.*, **2012**, *23*, 553-562.
(235) Purves, R. W.; Guevremont, R.; Day, S.; Pipich, C. W.; Matyjaszczyk, M. S., *Rev. Sci. Instrum.*, **1998**, *69*, 4094-4105.
(236) Purves, R. W.; Guevremont, R., *Anal. Chem.*, **1999**, *71*, 2346-2357.
(237) Xuan, Y.; Creese, A. J.; Horner, J. A.; Cooper, H. J., *Rapid Commun. Mass Spectrom.*, **2009**, *23*, 1963-1969.
(238) Saba, J.; Bonneil, E.; Pomiles, C.; Eng, K.; Thibault, P., *J. Proteome Res.*, **2009**, *8*, 3355-3366.
(239) Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L., *Nucleic Acids Res.*, **2007**, *35*, D301-D303.
(240) Shvartsburg, A. A.; Jarrold, M. F., *Chem. Phys. Lett.*, **1996**, *261*, 86-91.
(241) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., *J. Phys. Chem. A*, **1996**, *100*, 16082-16086.
(242) Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., *J. Chem. Phys.*, **1998**, *108*, 2416-2423.
(243) Wyttenbach, T.; Helden, G. v.; Batka Jr, J. J.; Carlat, D.; Bowers, M. T., *J. Am. Soc. Mass Spectrom.*, **1997**, *8*, 275-282.
(244) Clemmer, D. E.; Jarrold, M. F., *J. Mass. Spectrom.*, **1997**, *32*, 577-592.
(245) Bleiholder, C.; Contreras, S.; Bowers, M. T., *Int. J. Mass. Spectrom.*, **2013**, *354–355*, 275-280.
(246) Bleiholder, C.; Wyttenbach, T.; Bowers, M. T., *Int. J. Mass. Spectrom.*, **2011**, *308*, 1-10.
(247) Anderson, S. E.; Bleiholder, C.; Brocker, E. R.; Stang, P. J.; Bowers, M. T., *Int. J. Mass. Spectrom.*, **2012**.
(248) Bleiholder, C.; Contreras, S.; Do, T. D.; Bowers, M. T., *Int. J. Mass. Spectrom.*, **2013**, *345–347*, 89-96.
(249) Kelly, S. M.; Price, N. C., *Curr. Protein. Pept. Sci.*, **2000**, *1*, 349-384.
(250) Whitmore, L.; Wallace, B. A., *Biopolymers*, **2008**, *89*, 392-400.
(251) Whitmore, L.; Wallace, B. A., *Nucleic Acids Res.*, **2004**, *32*, W668-W673.
(252) Weakley, B. S. *Biological transmission electron microscopy*; Second edition ed.; Churchill Livingstone: Edinburgh, 1981.
(253) Jan K. G. Dhont; Gerhard Gompper; Peter R. Lang; Dieter Richter; Marisol Ripoll; Dieter Willbold; Zorn, R. *Lecture notes of the 42nd IFF Spring School 2011; Macromolecular Systems in Soft and Living Matter*; Forschungszentrum Julich GmbH: Julich, 2011; Vol. 20.
(254) Cohlberg, J. A.; Li, J.; Uversky, V. N.; Fink, A. L., *Biochemistry*, **2002**, *41*, 1502-1511.
(255) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M., *Nature*, **2002**, *416*, 507-511.
(256) Leavitt, S.; Freire, E., *Curr. Opin. Struct. Biol.*, **2001**, *11*, 560-566.

2

Experimental

This chapter contains descriptions of the instrumentation used for data collection, including typical instrumental operating conditions and schematic representations. An example of data acquisition and analysis for a drift tube ion mobility mass spectrometry experiment is also shown here.

2.1 Reagents

Liquid chromatography-mass spectrometry grade methanol and acetonitrile were purchased from Sigma Aldrich Company Ltd (Dorset, UK). High purity water, in the resistivity range of 18.0-18.2 MΩ·cm, was obtained from an Arium 611 water purification system (Satorius, Göttingen, Germany). Ammonium acetate used for buffers and ammonia solution used for pH adjustment were purchased from Fisher Scientific (Loughborough, UK). Formic acid and hydrochloric acid used for pH adjustment were purchased from VWR International Ltd (Leicestershire, UK). Sodium hydroxide used for pH adjustment was purchased from Sigma Aldrich Company Ltd (Dorset, UK). *Meta*-nitrobenzyl alcohol (*m*-NBA) used in 'supercharging' experiments was purchased from Sigma Aldrich Company Ltd (Dorset, UK). Sodium iodide, used in calibration of the Q-TOF instruments was purchased from Sigma Aldrich Company Ltd (Dorset, UK) and prepared in 50:50 methanol:water. All peptide and protein sample sources and preparation are discussed in the relevant chapters.

Solution pH measurements were taken using a Jenway 2505 pH meter (Jenway scientific equipment, Essex, UK), reading to 0.01 units. All sample solids were weighed to an accuracy of ± 0.0001 g on either a AC100 (Mettler Toledoscales, Columbus, OH, USA) or a AZ 124 (Satorius, Göttingen, Germany) balance.

2.2 Mass Spectrometry

All mass spectrometry (MS) experiments, excluding electron capture dissociation (ECD) experiments (see Section 2.5), were performed on quadruple time-of-flight (Q-TOF) mass spectrometers using nano-electrospray ionisation (n-ESI). Two Q-TOF instruments were utilised for the experiments presented in the following chapters, a Q-TOF II (Figure 2.1) and a Q-TOF Ultima API, both manufactured by Waters (Manchester, UK). The principles and operation of both instruments are similar and hence will be discussed as one in the following sections.

2.2.1 Nano electrospray ionisation

Samples were ionised using n-ESI in positive ionisation mode. The n-ESI tips were made in-house from thin-walled glass capillaries with an internal diameter of 0.5-0.9 mm (World Precision Instruments, Sarasota, FL, USA) using a Flaming/Brown micropipette puller (Sutter Instrument Company, Novato, CA, USA), yielding fine tips with diameters typically

in the range of 10-30 µm. The n-ESI tips were filled with sample using gel micro-loading tips (Eppendorf, Hamburg, Germany). The positive voltage required to ionise the sample was applied *via* a thin (0.125 mm diameter) platinum wire (Goodfellow, Huntingdon, UK). The voltage applied to the sample was typically in the range 1.2-2.0 kV, with the precise voltage applied depending on the sample solution composition, and often in the case of the glycosaminoglycan (GAG) binding studies the extent of aggregation observed. However, unless otherwise stated a capillary voltage of 1.6 kV was applied in all studies presented here.

2.2.2 An ion's journey in a Q-ToF: ion transfer, mass analysis and detection

In this section the journey of the ions from generation to detection in a Q-TOF mass spectrometer is described, Figure 2.1 highlights all major components of the Q-TOF II.

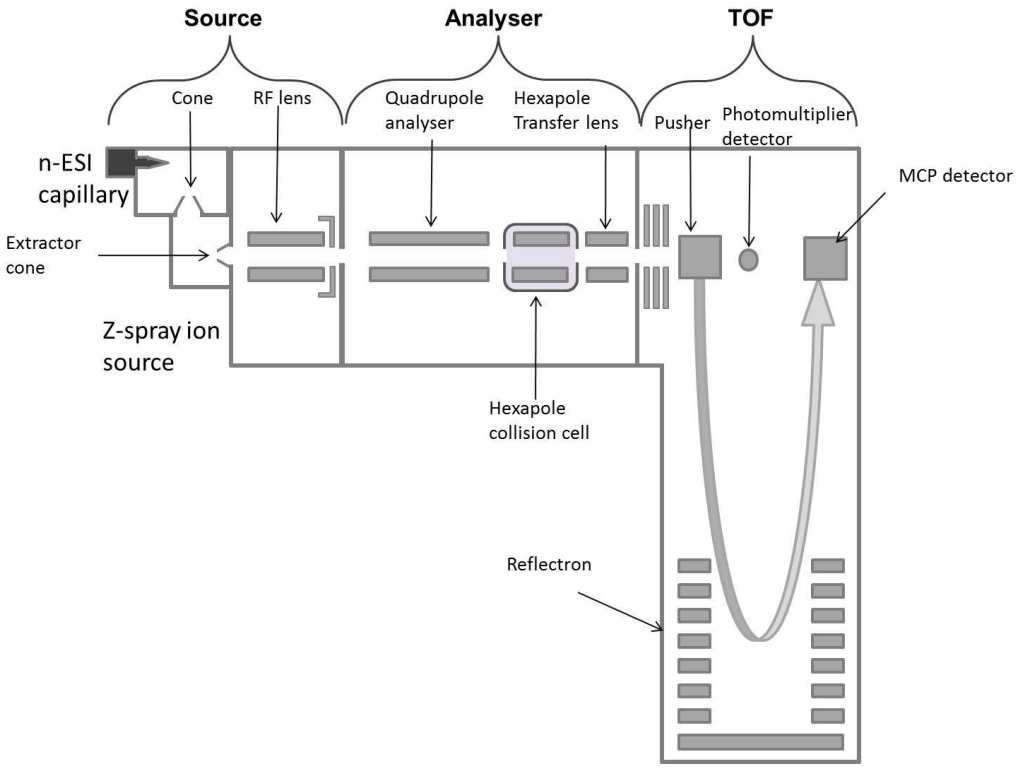

Figure 2.1: Schematic representation of the Waters Q-TOF II instrument. All major components are labelled and the ion path in the TOF is shown with an arrow.

The pressure of both Q-TOF instruments used here is controlled and maintained *via* a 250 or 255 L/s turbomolecular pump (Edwards, West Sussex, UK) in each of the differential pumping regions labelled as source, analyser and TOF in Figure 2.1. Source pumping and backing of the turbomolecular pumps is through two rotary pumps (Edwards, West Sussex, UK). The source pressure can be elevated through the partial closing of a SP 25 Speedivalve

(Edwards, West Sussex, UK) placed in the source pumping line between the source block and backing rotary pump. This is advantageous because an elevated source pressure can aid desolvation of protein complexes or aggregates, such as those observed in the GAG binding studies (Chapters 5 and 6). The source pressure was optimised for each system studied and typical values for each are shown in Table 2.1.

After generation by n-ESI, ions enter the mass spectrometer via a 'Z-spray' source. In this geometry the plume of ions formed is directed perpendicularly through the first sampling orifice (sample cone), and then through the second orifice (extractor cone) which is perpendicular to the first. The orientation of the Z-spray source allows greater sensitivity to be achieved, as ions, guided by decreasing voltages, enter the instrument, whilst limiting transmission of neutral species such as uncharged proteins and contaminants. In order to aid desolvation, and hence increase transmission, the source block was held at an elevated temperature, 80 °C unless otherwise stated.

Once ions have passed through the extractor cone they are then guided through an RF lens (transfer hexapole) to the analyser region. The typical pressure in this region was tuned for each system studied and is shown in Table 2.1.

System	***Instrument***	***Source pressure/ mbar***	***Analyser pressure/ mbar***
c-MYC:MAX	Q-TOF II	~1.8	2.3-3.5 x 10^{-5}
Lymphotactin	Q-TOF Ultima	~2.8-4.2 x 10^{-1}	4.3-5.2 x 10^{-5}
Human β-defensins	Q-TOF Ultima	~2.7-4.0 x 10^{-1}	4.3-5.2 x 10^{-5}

Table 2.1: Typical source and analyser pressure ranges used for each system studied and the instrument used.

Upon reaching the analyser region, the ions first pass through a quadrupole analyser, which can be operated in MS or MS/MS mode. In MS mode the quadrupole acts as a mass filter to allow all ions of a given *m/z* range to pass through and continue travelling the length of the instrument, through the hexapole collision cell and transfer lens to the TOF, whilst in MS/MS mode the quadrupole is used to select and isolate ions of a given *m/z* for subsequent analysis by collision induced dissociation (CID). In CID, ions of a preselected *m/z* are isolated in the quadrupole and transferred into the collision cell, which is filled with argon (BOC Speciality Gases Ltd, Guildford, UK). The kinetic energy with which these ions enter the collision cell can be increased, typically from 4-10 eV up to a maximum of 200 eV. Upon entering the collision cell ions undergo multiple collisions with the neutral argon gas.

Hence, increasing the kinetic energy with which ions enter the cell increases the energy of these collisions and facilitates ion fragmentation. After ions have exited the collision cell, either after fragmentation in MS/MS mode or passing straight through in MS mode, they travel through a final hexapole transfer lens before entering the TOF.

Ions enter the TOF through a number of lenses designed to focus and accelerate the ion beam into the time-of-flight analyser. The pusher then accelerates the beam orthogonally into the TOF tube until it reaches the retarding field of the reflectron, where ions are refocused and reflected back to the microchannel plate (MCP) detector. Each ion enters the TOF with the same kinetic energy and hence the time an ion takes to reach the detector is proportional only to its *m/z* and inversely proportional to its velocity. The signals detected by the MCP detector are transferred to a 4 GHz time-to-digitial convertor (TDC) which generates a mass spectrum by storing the arrival times of ions in a histogram memory. The output of the TDC is then processed and viewed as a mass spectrum in MassLynx^{TM} version 4.1 software, (Waters, Manchester, UK).

2.2.3 Calibration

The Q-TOF mass spectrometers were calibrated at the start of each day of experiments, through external mass calibration, using a calibration solution of 2 mg/mL sodium iodide prepared in 50:50 methanol:water. Sodium iodide solution produces a series of salt cluster ions of the type $Na_{n+1}I_n$ where $n \geq 0$, which span the range of *m/z* used throughout the experiments.

2.2.4 Typical Q-ToF II Settings

All c-MYC:MAX MS studies (Chapter 3) were performed on the Q-TOF II (Waters, Manchester, UK), typical operating settings of which are shown in Table 2.2.

2.2.5 Typical Q-ToF Ultima Settings

Ltn equilibria (Chapter 4), Ltn GAG binding (Chapter 5) studies and defensin GAG biding (Chapter 6) studies were performed on a Q-TOF Ultima API (Waters, Manchester, UK), the typical operating settings of which are shown in Table 2.3.

Instrument parameter	*Experimental Setting*
Capillary (kV)	1.6 - 1.8
Cone	23 - 25
Extractor	1
RF Lens	0.55
Source Temp (°C)	80
LM Resolution	5
HM Resolution	5
Collision Energy	4.9
Ion Energy	1.2
Steering	0
Entrance	70
Pre-filter	6
Transport	3.7
Aperture2	12
Acceleration	200
Focus	0
Tube Lens	100
Offset1	-0.5
Offset2	0
Pusher	980
TOF (kV)	9.1
Reflectron	34.66
Pusher Cycle Time (µs)	100 - 125
Multiplier	650
MCP	2000

Table 2.2: Typical experimental operating settings used for all experiments performed on the Q-TOF II. All values are in volts unless otherwise stated.

2.2.6 Q-TOF Ultima high-mass upgrade

A number of modifications were made to the Q-TOF Ultima by MS Vision (Almere, Netherlands) to enable the more efficient transmission of high mass species. The main modifications made are highlighted in this section and include an additional Speedivalve (SP10K, Edwards, West Sussex, UK) placed in the source backing line, enabling fine adjustment of the source pressure. Typical source pressure used post-modification was in the range 2.7-4.3 x 10^{-1} mbar. A pressure sleeve was fitted around the first few inches of the hexapole, increasing the pressure of this region, and the quadrupole mass range was increased to 30, 734 *m/z*. The apertures on the collision cell were reduced from 2.25 mm to 1.5 mm and the capillary supplying the collision gas was changed from one with an internal diameter of 75 µm to one with an internal diameter of 100 µm, the combination of these two alterations leading to increased pressure within the collision cell. Post-modification the

typical collision cell pressure used for experiments reported here was ~2.0 x 10^{-2} mbar. Additionally, the maximum collision energy was increased from 200 eV to 400 eV. Finally, the TOF pusher pulse was increased from 9 to 18 µs.

Instrument parameter	***Experimental Setting***	
	Pre upgrade	*Post upgrade*
Capillary (kV)	1.6 – 1.8	1.6
Cone	60 – 100	60
RF Lens1 Energy	1.7	2
Aperture 1	1 – 2.2	2
RF Lens2 Energy	0	0
Source Temp (°C)	80	80
LM Resolution	5	5
HM Resolution	5	5
Collision Energy	10	10
Ion Energy	1.7	1
Steering	0	-1.1
Entrance	75	65
Pre-filter	5.6	4
Transport	2.4	5
Aperture3	15.1	15
Acceleration	200	200
Focus	0	0
Tube Lens	90	60
Offset1	-0.7	0
Offset2	0	0
Pusher	975	980
TOF (kV)	9.1	9.1
Reflectron	35.52	35
Pusher Cycle Time (µs)	90 – 100	100 – 180
Multiplier	650	650
MCP	1800 – 2400	2400

Table 2.3: Typical experimental operating settings used for all experiments performed on the Q-TOF Ultima API, pre and post high mass upgrade. All values are in volts unless otherwise stated.

2.2.7 Typical Q-ToF Ultima Settings post high-mass upgrade

A number of experiments presented in this thesis were performed post high-mass upgrade including experiments to study lymphotactin and glycosaminoglycan binding (see Chapter 5 section 5.2) and a number of experiments to study the interactions of human β-defensins with glycosaminoglycans and each other (see Chapter 6 section 6.2). Typically instrumental parameters for all studies carried out post high-mass upgrade can be found in Table 2.3.

2.3 Ion mobility-mass spectrometry

All ion mobility experiments were performed on an in-house modified Q-TOF (Micromass UK Ltd, Manchester, UK), adapted in order to carry out separations based on an ion's mobility, and to enable the temperature-dependent collision cross sections (CCS) to be determined. The Q-TOF is similar to those described in detail in section 2.2.2. The main modifications made to this instrument, and its operation, are described in the following sections. In all experiments data was acquired using positive ionisation mode *via* n-ESI, as described in section 2.2.1. For details on sample preparation for all drift tube ion mobility-mass spectrometry (DT IM-MS) experiments please see the relevant chapters.

2.3.1 Instrument description: drift tube ion mobility-mass spectrometer

The main modification made to the Q-TOF instrument to enable the determination of an ions CCS was the inclusion of a 5.1 cm copper drift cell, and supplementary ion optics[1]. The drift cell is situated post-source, before the quadrupole analyser (Figure 2.2). For use the cell is filled with high purity (99.999 %) helium (BOC Speciality Gases Ltd, Guildford, UK), to a pressure of 3.8-4.0 Torr at a temperature of ~300 K (27 °C). The drift cell is situated within the drift chamber to which a 500 L/s TMH520 turbomolecular pump (Pfeiffer Vacuum Ltd, Newport Pagenll, UK) is fitted, which is backed by an dual-stage rotary pump (Edwards Vacuum, Crawley, UK), to address the extra load conferred by this chamber and by the buffer gas. The DT IM-MS instrument also contains the three 250 L/s turbomolecular pumps present on a standard Q-TOF; situated at the source, analyser and TOF differential pumping regions.

In order to have increased control over the pressure in the source region and to allow higher source pressures to be achieved, the DT IM-MS instrument was modified to include a speedivalve (Edwards, West Sussex, UK) placed in the source pumping line between the source block and the backing rotary pump, and an argon inlet into the source block (Figure 2.2). The pressure of the source was tuned for each sample, every day, to enable maximum ion transmission. Typical operating pressures for the DT IM-MS instrument are as follows; source ~1.3-4.2 x 10^{-1} mbar, drift chamber ~4.1-5.0 10^{-3} mbar, analyser ~2.0-3.1 x 10^{-6} mbar and TOF ~1.3-2.9 x 10^{-7} mbar.

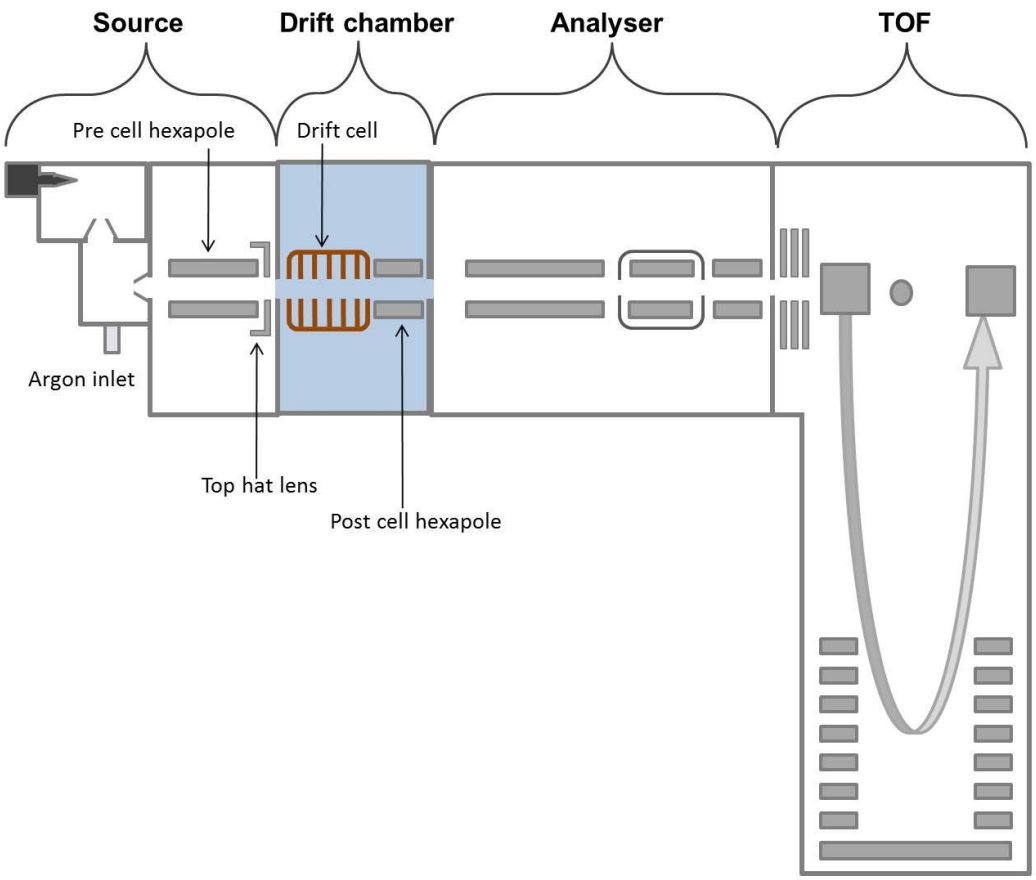

Figure 2.2: Schematic representation of the drift tube ion mobility mass spectrometer. Components of the Q-TOF are identical to those in Figure 2.1.

The DT IM-MS instrument can be operated either in MS or IM-MS mode. When operated in MS mode, the continuous beam of ions generated by n-ESI passes into the drift cell before travelling through the rest of the instrument. In IM-MS mode, ions are introduced into the drift cell as discrete pulses to enable the separation of analytes by their mobility. In order to create these pulses of ions the potential on the top hat lens is raised, causing accumulation of ions in the pre-cell hexapole. This trapping voltage is then lowered at regular intervals, for 40 µs, to inject a discrete pulse of ions into the drift cell.

The pulse of ions into the drift cell is synchronised with the TOF pulse, to allow recording of the mobility measurements, with one pulse into the drift cell (*i.e* one mobility separation) being sampled over 200 TOF pulses. In order to synchronise these pulses the frequency at which ions are pulsed into the drift cell is set, dependent on the TOF pusher period, whilst the TOF pusher period is itself dependent on the *m/*z range considered and hence sample being studied. The TOF pusher cycle time was in the range of 90-130 µs for all experiments reported here. The frequency (*f*) of drift cell pulse is calculated by $f = 1/t$, where time is the TOF pulse time (pusher cycle time) multiplied by the number of TOF pulses per ion mobility

separation (200). For example when a TOF pusher time of 130 µs is used, f = 1/(200*130 µs) = 38.5 Hz. This frequency is set and applied using a DG 535 digital delay/pulse generator (Stanford research systems, Sunnyvale, CA, USA).

Once the pulse of ions has been introduced, ions traverse the length of the drift cell due to a weak, static electric field applied to the cell, known as the drift voltage. The potential of this drift voltage is varied over the acquisition procedure, as discussed in more detail in Sections 2.3.2 and 2.3.3.

Upon exiting the drift cell ions are refocused by passing through a short post-cell hexapole, before travelling through the rest of the instrument to be detected by the MCPs, enabling the output of the ion mobility experiment to be recorded; this is known as the ion arrival time distribution (ATD). As the drift and TOF pulses are synchronised, such that one pulse into the drift cell is sampled over 200 TOF pulses, this means one ATD is generated for every 200 MS scans. A typical experiment will collect 11 ATDs and sum these to yield greater signal intensity. For samples with particularly low signal intensity, such as highly aggregated samples in GAG binding studies, 20 ATDs were acquired and summed for analysis.

2.3.2 Typical DT IM-MS settings

Tuning of the DT IM-MS instrument is controlled in two ways; Q-TOF mass spectrometer settings are controlled through a modified version of MassLynx^{TM} version 4.1 (Waters, Manchester, UK), whilst the drift cell and additional ion optics are controlled by an external power supply. The typical operating settings as controlled by MassLynx^{TM} for the modified Q-TOF used for all DT IM-MS experiments presented in the following chapters are shown in Table 2.4.

As mentioned above, the DT IM-MS instrument contains additional ion optics before and after the drift cell (Figure 2.3), these are controlled by an external power supply constructed in-house. The drift voltage (*i.e* the difference between C_1 and C_2, see Figure 2.3B) is altered stepwise to perform measurements at different drift fields, typically from 60 to 20 V. Voltages supplied to the pre and post cell lens are tuned to enable maximum transmission of ions. The ranges of settings for these lenses for all experiments shown here are given in Table 2.5. It is important to note that C_1, C_2, H_2 and TH_2 float upon the hexapole collision cell voltage, set in MassLynx^{TM}. Additionally, the pre-cell lens (H_1, TH_1 and L_1-L_3) are referenced to the cell body (C_1) and the post-cell lens (L_4) is referenced to the end-cap (C_2).

Instrument parameter	*Experimental settings*
Capillary (kV)	1.6 – 1.8
Cone	35 – 120
Extractor	130 – 110*
RF Lens	2 – 2.43
Source Temp (°C)	80‡
LM Resolution	4.4 – 5
HM Resolution	4.3 – 5
Collision Energy	0.9 – 4
Ion Energy	0.5 – 2.1
Steering	(-0.64) – (-2)
Entrance	40.8 – 49.3
Pre-filter	5 – 8.8
Transport	2.9 – 3
Aperture2	3.4 – 12.6
Acceleration	200
Focus	0 – 3
Tube Lens	90 – 100
Guard	44.1 – 50.2
TOF (kV)	7.2
Reflectron	35
Pusher Cycle Time (µs)	90 – 130
Multiplier	650
MCP	2400-3000

Table 2.4: Typical experimental operating settings used for all experiments on the DT IM-MS instrument. All values are in volts unless otherwise stated. *Extractor voltage is given when the drift voltage is set to 60 V, during DT IM-MS acquisition extractor is lowered by the same value as the drift voltage. ‡ For Ltn:GAG binding studies a source temperature of 100 °C was used to aid desolvation of these highly aggregating systems.

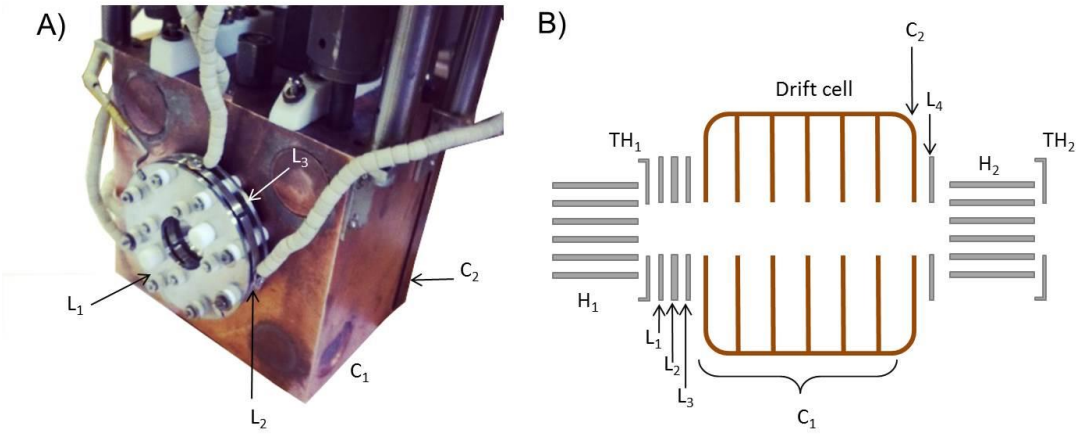

Figure 2.3: A) Photograph of the drift cell in the DT IM-MS instrument B) Schematic representation of the drift cell and lens stack, with lens labelled. H_1 and H_2 are the pre and post-cell hexapoles respectively. TH_1 and TH_2 are the pre and post-cell top hat lenses. Pre-cell lens (L_1-L_3) and post-cell lens (L_4) are also labelled. C_1 is the cell body and C_2 the end-cap.

Lens	Applied voltage range / V	Output voltage
TH_1	0 – 33	C_1 + CV +TH_1
H_1	33 – 41	C_1 + CV + H_1
L_1	(-1) – (-276)	C_1 + CV + L_1
L_2	0 – (-112)	C_1 + CV + L_2
L_3	3 – 32	C_1 + CV + L_3
C_1	70 – 30*	C_1 + CV
C_2	10	C_2 + CV
L_4	0 – (-25)	C_2 + CV + L_4
H_2	0	H_2 + CV

Table 2.5: The drift cell lens applied voltage range used in experiments presented here and the corresponding output voltages. *The voltage applied to C_1 is altered throughout the course of the mobility experiments, affording the different drift voltages.

2.3.3 The experimental workflow: an example

In order to demonstrate the typical acquisition and analysis procedure for a DT IM-MS experiment an example is given here for a plant cyclotide known as kalata B1, see Figure 2.4.

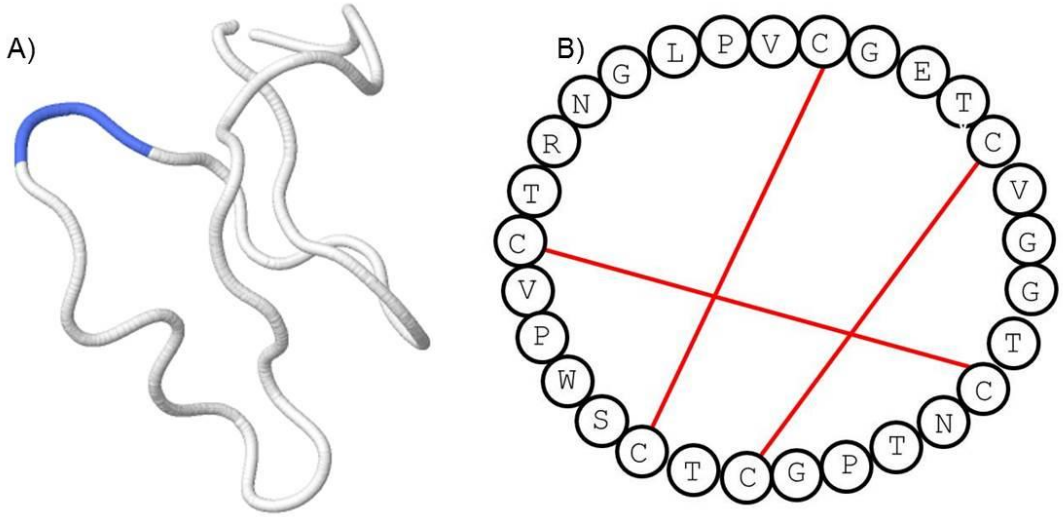

Figure 2.4: Kalata B1 A) Structure solved by NMR (PDB 1JJZ), B) Cartoon representation showing sequence of amino acids, with disulfide bonds shown by red lines. Cyclic kalata B1 used in this study was prepared by the MacMillan research group using bacterial expression followed by cyclisation[2].

The first stage of any DT IM-MS experiment involves the tuning of the source pressure and drift cell optics for the sample of interest. This is performed with the instrument in MS mode and following tuning a MS spectrum is acquired for reference. Then in IM-MS mode, at a drift voltage of 60 V, the total arrival time distribution is collected. Typically 11 ATDs are collected (Figure 2.5A) and summed for analysis; however, as discussed in section 2.3.1 the number of ATDs collected can be increased for samples of weak intensity. The potential applied to C_1 is then lowered, lowering the drift voltage, and the ATDs are collected again at this lower drift voltage. ATDs are collected over eight drift voltages which are as follows;

60, 50, 45, 40, 35, 30, 25 and 20 V. The temperature and pressure at the start and end of each acquisition is recorded. These values are then averaged and reported as the temperature and pressure for each drift voltage acquisition.

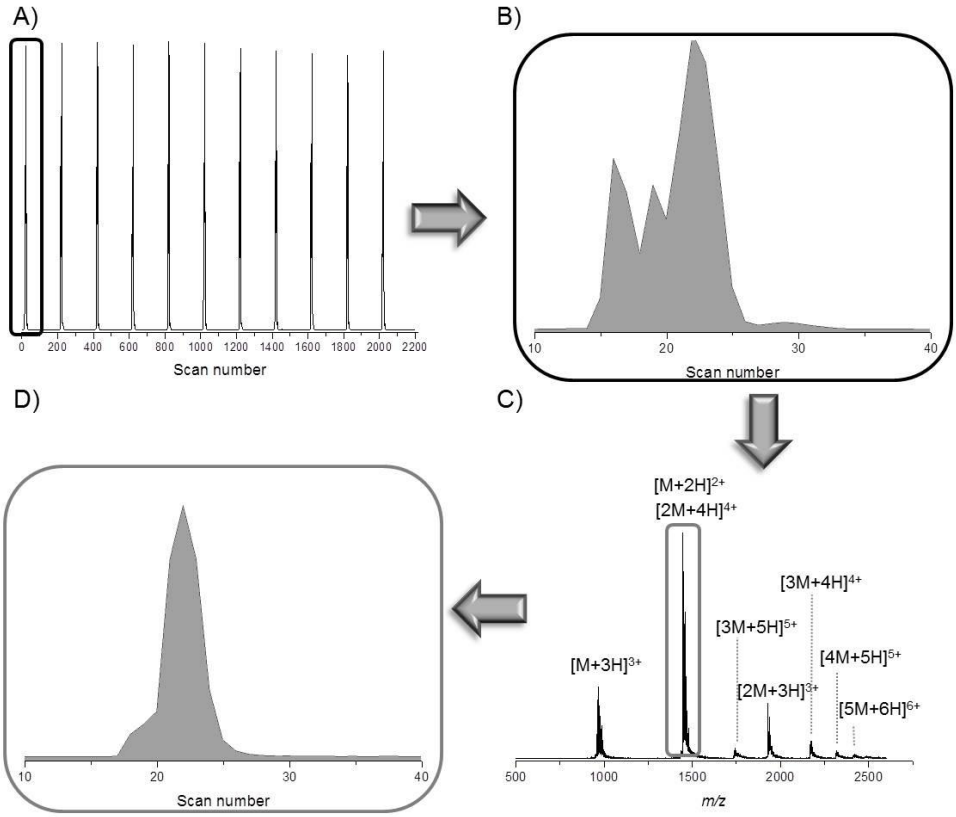

Figure 2.5: DT IM-MS experimental workflow, an example using kalata B1. A) The 11 total ATDs obtained at a drift voltage of 45 V, B) The region highlighted in A magnified, to show a single total ATD at a drift voltage of 45 V C) Full mass spectrum of all ions arriving in the single mobility separation shown in B, where M represents a single Kalata B1 peptide, and D) the extracted ion ATD for the peak highlighted in C, at *m/z* 1447 at a drift voltage of 45 V.

Using MassLynx^{TM} version 4.1, the full mass spectrum of all ions arriving at the detector following a single ion mobility experiment (Figure 2.5B) can be generated by combining over all scans under the peak (Figure 2.5C). The mass spectrum of cyclic oxidised peptide kalata B1 obtained in this way shows that this peptide exists in a number of oligomeric states presenting from monomeric (M) species up to pentameric (5M) species. The MassLynx^{TM} software enables an extracted ion ATD to be obtained for each peak in the mass spectrum, achieved by combining over the *m/z* range of interest. For the purpose of this example analysis will focus on the peak at *m/z* 1447, which could be due to two mass-coincident species namely the $[M+2H]^{2+}$ and $[2M+4H]^{4+}$. The dimeric species $[2M+4H]^{4+}$ has double the mass but also double the charge and hence will exist at the same *m/z* ratio as the

monomeric species, $[M+2H]^{2+}$. The extracted ion ATD of the peak at *m/z* 1447 (Figure 2.5D) does indeed show two distinct peaks, indicating the presence of two species.

In an extracted ion ATD the presence of two peaks can be due to either species of the same *m/z* but of different oligomeric orders or the presence of multiple conformations, both of which can be separated out based on their mobility. These two phenomena can be distinguished from each other by considering the relative ratio of the two species in the extracted ion ATD at multiple drift voltages. The relative ratio between the two can remain constant, suggesting it is two conformers of a species. Alternatively, the relative ratio may change with drift voltage with the earlier arriving conformer increasing in relative intensity with decreasing drift voltage, which suggests that the two species are due to ions of differing oligomeric orders. In this case the increase in the relative proportion of the higher order oligomer is due to the fact that it is more highly charged, with higher charged species being better focused by the confining voltages applied and less susceptible to radial diffusion. These effects become more pronounced at lower drift voltages as ions spend longer travelling the length of the drift cell, due to the decreased electric potential. In the case of the ion at *m/z* 1447 for kalata B1, these two species are due to a monomeric and dimeric species, as determined from the extracted ion ATDs over the eight drift voltages and from the mass spectrum, in which we can clearly see the presence of dimeric species, namely the $[2M+3H]^{3+}$ species at *m/z* 1929.

Once the extracted ion ATDs are generated, the intensity is summed over the total number of pulses in a Microsoft Excel (Microsoft, Redmond, WA, USA) spreadsheet. The summed extracted ion ATDs are then input into Origin 8.5.1 or Origin 9 (OriginLab, Northampton, MA, USA), which is used to determine the average arrival scan number. Peaks in the ATD were fitted using an iterative Gaussian function in Origin (Figure 2.6A), which considers the following parameters; peak centre, width, area and y-offset. In all cases a y-offset of zero was applied and all other parameters were determined through iterative fitting for each extracted ion ATD, at each drift voltage. The Gaussian function enables the centre of the peak to be determined, which is used as the average scan number for each discrete peak.

The same procedure is applied to extracted ion ATDs containing only a single resolvable conformational family, in this case with only one Guassian peak being considered and its centre used as the average scan number. The number of conformational families and hence number of peaks fitted for each ATD was determined by first visually considering the ATD at each drift voltage, in addition to attention to the quality of fit in all cases where an R^2 value of ≥ 0.9990 was used to indicate a fit of good quality. As the drift voltage decreases the

time ions spend traveling in the drift region increases and hence distinct conformational families or oligomers can become resolvable; presenting either as distinct peaks, shoulders or tails. If multiple conformational families are observed and an average arrival scan number determined over five or more drift voltages then the CCS is reported. In cases where multiple conformational families are observed, but not resolvable, the presence of an additional conformational family is noted and reported, however, CCS cannot be accurately determined. For all samples studied in this thesis, where multiple conformations were observed, both species were clearly resolvable in one or more repeats and such species are clearly identified in the relevant Chapters and Appendices.

In all cases the minimum number of peaks to obtain a fit of good quality, were considered. It is important to note that due to the resolution of the instrument in some cases similar conformational families will not be resolvable, in such cases the width of the ATD also provides information on the conformational diversity of the species of interest. A protein may exist in a wide range of similar conformations which, if there is a small variance in the different CCS, may not be resolvable on this instrumentation. The width of the ATD is therefore considered for each protein over all charge states and drift voltages, as a measure of conformational diversity. For all samples studied in this thesis, the width of the ATD either indicated a single conformational family, in which case a single average CCS is reported for each charge state; or multiple resolvable conformations, in which case the CCS of each conformational family is reported.

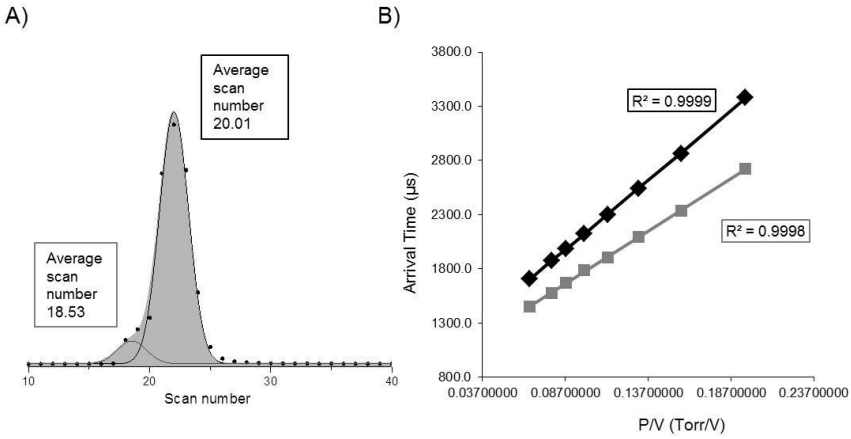

Figure 2.6: DT IM-MS analysis for cyclic oxidised kalata b1 peak at *m/z* 1447 A) Extracted ion ATD at drift voltage of 45 V, peaks fitted using the Guassian function in Origin 9 with average scan number as determined from the centre of each peak labelled. B) Plot of arrival time versus P/V for each of the eight drift voltages for monomeric species (black) and dimeric (grey), fitted with a line of best fit. Corresponding R^2 values are shown.

For CCS determination the Guassian fitting procedure is repeated for each ion at each drift voltage. The average scan numbers obtained are then converted to the average arrival time,

by multiplication with the MS pusher time. The average arrival time (t_a) of a particular ion corresponds to the length of time an ion spends travelling through the drift cell, the drift time (t_d), combined with the length of time the ion spends travelling through the rest of the mass spectrometer, known as the dead time (t_0), *i.e* $t_a = t_d + t_0$.

The principles of drift tube ion mobility-mass spectrometry are well established and it is known that the low field mobility (K) of an ion in a drift cell is inversely proportional to its drift time and the drift voltage (V) (Equation 2.1), where L is the length of the drift region. The low field region is defined as having a low ratio of electric field strength (E) to buffer gas number density (N). At low fields, ions have low velocities, which are independent of the field strength and therefore the motion of the ions can be described in more simplistic terms.

$$K = \frac{L^2}{t_d V} \qquad [2.1]$$

The mobility of an ion, however, is also dependent on the temperature (T) and pressure (P) at which the experiment is carried out, hence in order to normalise for differences in temperature and pressure the reduced mobility is most often reported (K_0), see Equation 2.2.

$$K = K_0 \frac{P_0 T}{T_0 P} \qquad [2.2]$$

Where P_0 and T_0 are the reduced temperature (273.15 K) and pressure (760 Torr) respectively.

By combining Equation 2.1 and 2.2 (Equation 2.3) it is clear that a plot of average arrival time versus (P/V) for each ion will yield a linear plot with a gradient which is inversely proportional to the reduced mobility of that ion (Figure 2.6B) where P is the average measured experimental pressure for each drift voltage. The R^2 value of this line is used to determine the quality of the experimental data, with a value of ≥0.9990 equating to good quality, usable data and indicative of the fact that the experiments were performed within the low field region, which is imperative for the relationships described above to hold. Furthermore, the y-intercept of this plot is equal to the deadtime (t_0).

$$t_d = t_a - t_0 = \frac{L^2 T_0 P}{K_0 T P_0 V} \qquad [2.3]$$

Therefore, once the average scan numbers and corresponding arrival times have been calculated for an ion, over the eight drift voltages, it is then possible to plot these against P/V

and determine both the reduced mobility (K_0) and the dead time (t_0) of the ion. The rotationally averaged collision cross section (Ω) of the ion can then be determined from Equation 2.4.

$$\Omega = \frac{3ze}{16N}\left(\frac{2\pi}{\mu k_B T}\right)^{\frac{1}{2}}\frac{1}{K_0} \qquad [2.4]$$

Where, z is the nominal ion charge state, e is the elementary charge, N is the buffer gas density, μ is the reduced mass of the ion-neutral pair, T is the gas temperature and k_B is the Boltzmann constant.

In all samples studied, this process is repeated on three different days, yielding CCS measurements in triplicate. The average value is then reported and the error reported as the standard deviation between these three values. For kalata B1 the CCS of the $[M+2H]^{2+}$ and $[2M+4H]^{4+}$ ions were calculated, using the procedure outlined above, as 452 ± 4 and 686 ± 5 $Å^2$ respectively. Kalata B1 was chosen as an example here to highlight not only the data acquisition and analysis procedure for a DT IM-MS experiment but also the intrinsic ability of IM-MS to separate out ions of the same *m/z* but differing oligomeric orders. This feature of ion mobility mass spectrometry is particularly useful in the study of highly aggregating systems, as these species would not be distinguishable in a typical mass spectrometry experiment unless extremely high resolution instruments were used which, due to their often limited *m/z* transmission range, are not always applicable to the study of large proteins or protein complexes.

Another elegant feature of DT IM-MS is its ability to distinguish between different conformations of an ion. For example the Kalata B1 peptide can be produced in a cyclic oxidised or reduced form and a linear form, which is typically reduced. Although these species would have slightly differing masses we can distinguish them further based on their conformations using DT IM-MS, as has been shown previously for other cyclic and linear peptide systems[3]. The CCS of $[M+2H]^{2+}$ for cyclic oxidised, cyclic reduced and linear Kalata B1 were experimentally determined to be as follows; 452 ± 4, 463 ± 4 and 474 ± 4 $Å^2$ respectively. The cyclic oxidised form presents with the smallest CCS suggesting it is the most structurally constrained, as expected since this form contains three disulfide bonds. The cyclic reduced form is slightly larger, suggesting the absence of the disulfide bond enables it to adopt a larger conformation, and the linear form, as expected, adopts the largest conformation and hence largest CCS, as it is the least structurally constrained, containing no disulfide bonds and not being constrained through cyclisation.

2.4 Molecular modelling

DT IM-MS experiments provide information on the global fold of a protein or peptide at any given charge state, through determination of experimental CCS. In order to gain more detailed insight into the conformations adopted by a species it is possible to compare experimental CCS obtained to theoretical CCS obtained from molecular modelling. This is afforded either through the comparison from solved structure coordinates (from NMR or X-ray crystallography experiments) or from molecular dynamics (MD) simulations. In the following chapters, comparisons are made using both methods. The theoretical CCS of a structure, both from NMR coordinates or MD simulations, is determined using the program MOBCAL[4,5] which provides algorithms to determine the mobility or momentum transfer integral (rotationally averaged collision cross section). MOBCAL provides four methods to determine CCS; projection approximation (PA), exact hard sphere scattering model (EHSS), trajectory method (TM) and projection super approximation (PSA), as discussed in Chapter 1 section 1.5.4.4. In the work presented here only the TM values were used and reported. All molecular modelling was performed by Dr Massimiliano Porrini and brief details of the procedures used to determine the theoretical CCS are given in the relevant chapters. Unless otherwise stated, the ionisation state of the protein or peptide used for calculations was assigned using the *tleap* module in Amber, which uses an algorithm to automate assignment of protonated side chains, computing pK values of ionisable groups and adds in any missing hydrogen atoms to the structure in use[6,7].

2.5 Fourier transform ion cyclotron resonance mass spectrometry

All high resolution mass spectrometry and electron capture dissociation (ECD) experiments were performed on a 12 T Apex Ultra Qh FT-ICR mass spectrometer (Bruker Daltonik Gmbh, Bremen, Germany). The Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS) instrument was calibrated for each day of experiments using ESI-L, a low concentration tune mix from Agilent Technologies (Santa Clara, CA, USA).

This hybrid mass spectrometer contains a series of ion optics post-source and pre-ICR cell to guide ions into the high vacuum analyser cell. After sample ionisation, using either ESI or n-ESI, the ions are guided through a capillary and a series of ion funnels before passing through a multipole. The ions then pass through a series of focusing lenses before entering the quadrupole analyser, in which they can be mass isolated, they are then refocused in a

second series of focusing lens which guides the ions into the collision cell, where CID can be employed. Finally, ions pass through the last set of ion focusing lens before travelling along the transfer line into the ICR cell, in which ECD fragmentation occurs.

For analysis of monomeric wild type (WT) Lymphotactin (Ltn), sample ionisation was *via* ESI, with a capillary voltage of 4 kV. For all other studies n-ESI was performed using a NanoMate (Advion Biosciences, Ithaca, NY, USA) running in infusion mode and equipped with a HD_A_0 ESI chip (Advion biosciences, Ithaca, NY, USA) using a capillary voltage of 1.6 kV. For sample preparation details please see the relevant chapters (Chapters 4-6).

2.5.1 Electron capture dissociation

For ECD experiments the ECD current was 1.7 A, the ECD cathode was turned on and allowed to warm up for at least one hour before acquisition began. In order to maximise ion signal during isolation, for each peptide, protein or complex studied, each charge state was tuned individually, typical values for each system are shown in Table 2.6.

	WT Ltn	***WT 1-72 Ltn***	***CC3 Ltn***	***W55D Ltn***	***WT Ltn:Fx***	***HBD2***	***HBD2: Fx***	***HBD2: Fx + m-NBA***
ECD pulse length /s	0.006-0.01	0.02	0.015	0.015	0.015	0.02	0.015	0.01-0.02
Skimmer /v	60–65	50–90	40– 75	50 – 85	80–120	60	65–70	60
Time of flight / ms	1–1.1	2–2.2	1.0	0.9–1.1	2.0	1.2	1.2	1.2
Ion accumulation /s	1–10	1–4	0.2–10	0.5–8	5–10	0.1–0.6	2.0–6.0	0.8–1.2
Total scans	300	300	200	200	200	300	200	300

Table 2.6: Typical FT ICR MS operating parameters optimised for ECD studies and their settings for each system studied.

All data was reprocessed in DataAnalysis 4.0 (Bruker Daltonik Gmbh, Bremen, Germany). Peaks were identified and charge states and molecular weight determined using the SNAP algorithm in the DataAnalysis software with the following parameters; signal-to-noise ratio 0.0001, relative intensity threshold 0.0001 and quality factor threshold 0.001. ProSight PTM (Kelleher research group, Northwestern University, Evanston, IL, USA) was used to identify c' and $z^{\bullet}$-type fragments with a fragment tolerance set up to 50 ppm. Conventional ECD results in complementary c' and $z^{\bullet}$ fragment ion pairs (see Chapter 1 section 1.5.3.4), however, hydrogen transfer in long-lived intermediates can result in $c^{\bullet}$ and z' fragment

formation[8,9] and is an important consideration for compact protein species[10], such as those studied here. To enable consideration of species formed following hydrogen transfer all data was manually inspected following analysis by ProSight PTM. Internal (zIc) fragments were also identified manually.

Site specific fragment yields were calculated for monomeric fragmentation species as percentage values, considering that backbone cleavage to form c and z fragments can give a pair of complementary ions and internal cleavage can yield a complementary triad of ions. Considering this, 100% yield= 0.5[c] + 0.5[z] + 0.33[internal fragments] + [other products], where other products are reduced molecular ions and products from loss of small neutral species from the later. For fragmentation of dimeric species, site specific fragment yields were calculated as such 100% yield= 0.5[c] + 0.5[z] + 0.17[internal fragments] + [other products]. For fragments formed by dissociation of dimer to monomer followed by ECD fragmentation the site specific yields were calculated as such 100% yield= 0.25[c] + 0.25[z] + 0.17[internal fragments] + [other products], where other products also involves monomeric species arising from dimer dissociation. Low backbone fragmentation yields are observed, in all cases, due to the nature of ECD where the dominate species observed are due to electron capture, and charge reduction, without cleavage. In order to overcome the low fragmentation efficiency in ECD, and to increase fragmentation yields, 'activated ion' ECD is often implemented[11-13], which involves either infra-red irradiation or collisional activation. Activated ion ECD has the advantage of enabling higher fragmentation yields and greater sequence coverage to be obtained, however, can also promote unfolding of ions[13]. In all studies reported here ions were not activated prior to the introduction of electrons in ECD.

2.6 Circular dichroism spectroscopy

The secondary structure of the c-MYC-Zip, c-MYC-ZipΔDT and MAX-Zip peptides was assessed using circular dichrosim (CD) spectroscopy. The secondary structure present in mixed peptide solutions containing either c-MYC-Zip plus MAX or c-MYC-ZipΔDT plus MAX were also assessed, in the absence and presence of the ligand 10058-F4. Information on the sample preparation, including buffer strength, pH and concentrations used can be found in Chapter 3 section 3.2.

All CD spectra were recorded using a quartz cuvette with a 1 mm path-length (Hellma, Essex, UK) on a J-810 spectropolarimeter (Jasco, Essex, UK). Spectra were recorded at 25 °C, over the range 190-290 nm, at a rate of 10 nm min^{-1}. Five scans were acquired for each sample and averaged to give the final spectrum. Buffer or buffer plus ligand, in the

absence of peptide were used as reference, and the background was subtracted for such. The spectra were smoothed using the Savitzky-Golay algorithm, with a convolution width of 25. The secondary structure content was estimated using the CDSSTR algorithm on the DICHROWEB server[14-16].

2.7 Transmission electron microscopy: preparation of grids

All transmission electron microscopy (TEM) experiments were performed on a Philips CM120 Biotin transmission electron microscope (Philips, Eindhoven, Netherlands) with an operating voltage of -80 kV. TEM studies were performed on solutions of WT, WT 1-72, CC3 and W55D Ltn in the presence of Fondaparinux (Fx) and WT in the absence of Fx, for information on sample preparation see Chapter 5 section 5.2. Defensin TEM studies were performed on individual solutions of HBD2 and HBD3 in the presence of Fx and HBD3 in the absence of Fx. Further studies on HBD2 plus HBD3 in the presence of Fx were also performed. For information on defensin sample preparation for TEM see Chapter 6 section 6.2.

The preparation of TEM grids for analysis was as follows; 3 µL of sample was spotted onto a 3.05 mm, 200 mesh copper grid (Taab, Aldermaston, UK) coated with formvar and carbon and left for five minutes. The grid was then dried and rinsed once with 10 µL high purity water, after which the excess water was wicked off. The sample was then stained using 4 µL 1 % uranyl acetate (w/v) and left for 30 seconds before the excess stain was wicked off. The grid was then left to dry overnight.

2.8 Isothermal Titration Calorimtery

Isothermal titration calorimetry (ITC) experiments were performed on HBD2 plus Fx, to study the solution phase interaction of these systems. All ITC experiments were performed using a MicroCal Auto-ITC_{200} (GE Healthcare and Life Sciences, Little Chalfont, UK). For all ITC experiments the reference cell was filled with water and a reference voltage of 5 V was applied whilst the sample cell contained the HBD2 solution and Fx was automatically titrated in using the pipette. Samples were stored at 25 °C for at least 30 minutes prior to analysis and all ITC experiments were performed with the instrument set to 25 °C. The sample cell was stirred at a constant speed of 1000 revolutions per minute (RPM). For all experiments a total of 16 injections were performed with 180 seconds between each injection

to allow the system to re-equilibrate, the first injection was 0.5 µL and all subsequent injections were 2.49 µL in volume.

Two experiments were performed injecting the Fx into a HBD2 sample solution at different concentrations and two control experiments were performed, Fx was injected into a cell filled with buffer and buffer was injected into a cell containing HBD2. For all ITC experiments the buffer was 20 mM ammonium acetate, for sample preparation details see Chapter 6 section 6.2. Between each ITC experiment the pipette used to inject the Fx or buffer into the cell was cleaned with methanol and Decon^{TM} Contrad^{TM} detergent (Decon Laboratories, East Sussex, UK). All data was analysed in Origin 9.

2.9 References

(1) McCullough, B. J.; Kalapothakis, J.; Eastwood, H.; Kemper, P.; MacMillan, D.; Taylor, K.; Dorin, J.; Barran, P. E., *Anal. Chem.*, **2008**, *80*, 6336-6344.
(2) Cowper, B.; Craik, D. J.; Macmillan, D., *ChemBioChem*, **2013**, *14*, 809-812.
(3) Macmillan, D.; De Cecco, M.; Reynolds, N. L.; Santos, L. F. A.; Barran, P. E.; Dorin, J. R., *ChemBioChem*, **2011**, *12*, 2133-2136.
(4) Shvartsburg, A. A.; Jarrold, M. F., *Chem. Phys. Lett.*, **1996**, *261*, 86-91.
(5) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., *J. Phys. Chem. A*, **1996**, *100*, 16082-16086.
(6) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., *Nucleic Acids Res.*, **2005**, *33*, W368-371.
(7) Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A., *Proteins: Struct. Funct. Bioinf.*, **2006**, *63*, 928-938.
(8) O'Connor, P. B.; Lin, C.; Cournoyer, J. J.; Pittman, J. L.; Belyayev, M.; Budnik, B. A., *J. Am. Soc. Mass Spectrom.*, **2006**, *17*, 576-585.
(9) Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A., *J. Am. Soc. Mass Spectrom.*, **2007**, *18*, 113-120.
(10) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, U. A.; Tsybin, Y. O., *Chem. Soc. Rev.*, **2013**, *42*, 5014-5030.
(11) Horn, D. M.; Ge, Y.; McLafferty, F. W., *Anal. Chem.*, **2000**, *72*, 4778-4784.
(12) Horn, D. M.; Breuker, K.; Frank, A. J.; McLafferty, F. W., *J. Am. Chem. Soc.*, **2001**, *123*, 9792-9799.
(13) Breuker, K.; Oh, H. B.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W., *J. Am. Chem. Soc.*, **2002**, *124*, 6407-6420.
(14) Whitmore, L.; Wallace, B. A., *Nucleic Acids Res.*, **2004**, *32*, W668-W673.
(15) Whitmore, L.; Wallace, B. A., *Biopolymers*, **2008**, *89*, 392-400.
(16) Johnson, W. C., *Proteins-Structure Function and Genetics*, **1999**, *35*, 307-312.

3

Investigating disorder-to-order transitions:

c-MYC:MAX

Understanding the disorder-to-order transition that can occur upon the binding of a protein to its partner is essential for better understanding of both protein structure and function. This is particularly important for an aggressive oncoprotein such as c-MYC, which contains a disordered region that forms a leucine zipper upon binding its partner protein MAX, enabling it to carry out its function. In order to understand and potentially inhibit this interaction in initial studies it is more feasible to look at the regions directly involved in the interaction. The leucine zipper regions of c-MYC and MAX and the formation of their homo- and heterodimers have been examined using CD, MS and DT IM-MS. The effect on the heterodimer formation of adding a previously identified inhibitor (10058-F4) was also studied. Experimental DT IM-MS results were compared to results from MD studies in order to help elucidate the different dimeric conformations observed.

3.1 Introduction

Intrinsically disordered proteins (IDPs) are promiscuous proteins which have been shown to be functionally important whilst being notoriously difficult to study using traditional protein characterisation techniques[1-3]. Disorder-to-order transitions or coupled binding and folding are thought to be important pathways of IDPs, frequently associated with signalling events in the cell and mediating protein:protein interactions[4-6]. Interactions of IDPs with their targets facilitate binding with moderate affinities (as part of the binding free energy is expended upon folding of the IDP on binding) but high specificity, which is thought to be critically important for regulation and signal transduction[7]. Furthermore, IDPs have the ability to bind multiple partners, often adopting different folded forms upon binding[8]. Hence there are a number of different disorder-to-order transitions possible for such systems; a particularly common transition, however, is from disordered to helical[9]. Studying the conformations adopted upon disorder-to-order transitions and the mechanisms by which they occur is incredibly important for the understanding of these pathways. An elegant study by Keppel *et al*[10] presented the use of hydrogen/deuterium exchange and trypsin digest mass spectrometry to study the coupled folding and binding of two IDPs in a protein complex. Hydrogen/deuterium exchange is a method which can probe backbone flexibility by studying how quickly amide hydrogens can exchange with deuteriums when the protein is diluted in an excess of D_2O. Their study considered the IDP ACTR (activator of thyroid and retinoid receptors, NCOA3_HUMAN, residues 1018-1088) and its binding partner CBP, which is a molten globular protein. The results highlighted that regions involved in binding and/or folding could be readily identified due to increased protection in the complex, as compared to the individual species, and elegantly demonstrated the emerging power of mass spectrometry to study these disordered systems and their interactions.

IDPs are known to play crucial roles in many important diseases such as Parkinson's, Alzheimer's and many others, in addition to being implicated in cancer initiation, progression and metastasis[11,12]. Drug discovery efforts aimed at these highly dynamic systems have, however, been limited and they are still often considered 'undruggable' due to the difficulty in applying traditional drug discovery techniques, *i.e.* the design of a molecule that can 'fit' to a well-defined target site. Approaches to design drugs to target these important, flexible proteins are, however, on the rise, and have generally been aimed at regulating the protein:protein interactions; nonetheless the rational design of IDP ligands remains a challenging task[13-15]. Previous studies have suggested that binding of a small molecule to an IDP can stabilise its disordered state thereby exploiting the entropic gain of

disordered system to prevent target binding[16], as has been shown for the intrinsically disordered transcription factor c-MYC[17].

c-MYC is an oncoprotein and transcription factor, which can only carry out its full function following binding with its obligatory co-factor MAX[18,19] (see Chapter 1 section 1.2). Heterodimerisation occurs through the binding of a basic helix-loop-helix leucine zipper (b-HLH-Zip) domain in MYC to a similar region on MAX. The b-HLH-Zip regions of MYC and MAX are intrinsically disordered until they dimerise, when they form a parallel left-handed four-helix bundle with each monomer, composed of two α-helices, separated by a loop[20].

MYC is highly deregulated in an incredibly broad range of human cancers (Chapter 1 section 1.2.3), hence making it an attractive target for novel tumour specific cancer treatments[21]. Despite numerous different studies (targeting MYC at various levels[21]) to date no anti-MYC drugs are clinically available. One approach to targeting c-MYC is by preventing c-MYC:MAX heterodimerisation to form the leucine zipper complex (Figure 3.1A). The Prochownik and Metallo research groups have reported previously a number of c-MYC inhibitors with low μM affinity aimed at stabilizing the c-MYC disordered state and preventing heterodimerisation[17,22,23]. These initial studies have focused on peptide sequences of c-MYC and MAX as opposed to full length protein. One such report by Hammoudeh *et al*[17] identified three possible binding sites for a number of these ligands two of which lie within the b-HLH-Zip region, with the most potent c-MYC:MAX inhibitor 10058-F4, (Figure 3.1B) binding to a 7 amino acid stretch of c-MYC that overlaps with the leucine zipper region.

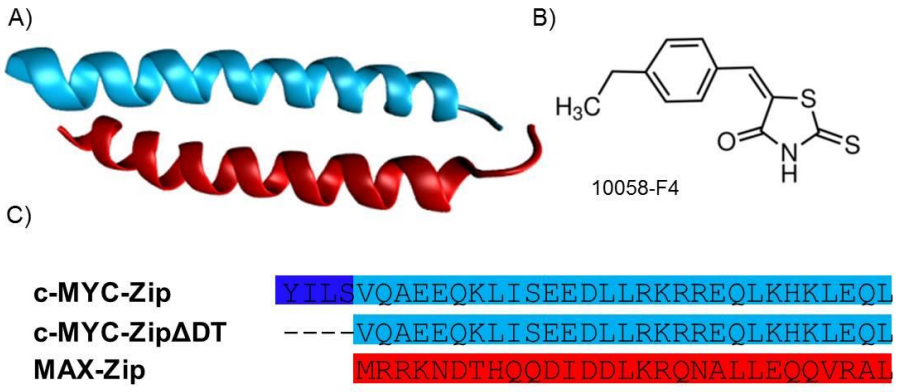

Figure 3.1: A) The NMR structure of the c-MYC:MAX leucine zipper (PDB file 1A93). B) Proposed c-MYC inhibitor 10058-F4. C) Peptides synthesized for this study: the leucine zipper region of MAX and c-MYC, along with a longer c-MYC peptide also containing the proposed 10058-F4 target site.

In this chapter, in order to study the interaction of 10058-F4 with c-MYC and any disruption in the formation of the c-MYC:MAX heterodimer, the leucine zipper regions of both proteins were synthesised, in addition a shorter c-MYC peptide which consists of the c-MYC leucine zipper region but does not contain the additional amino acids which make up the full drug target region (Figure 3.1C). The peptides used here are similar to those used in previous studies of c-MYC:MAX[17,22-24].

Due to the flexibility of IDPs, experimental approaches for characterising their structures and interactions with ligands are limited[25,8]. Traditional techniques for the study of protein structure, such as NMR spectroscopy or X-ray crystallography have produced some undoubtedly remarkable findings for structured proteins[26,27], however, their application to IDPs is often limited due to the large quantities of material needed, difficulties in distinguishing between species with extensive conformational heterogeneity and difficulty in producing high quality crystals. Due to its ability to study dynamic systems presenting with multiple conformations, coupled with the low sample quantities required, mass spectrometry is becoming an increasingly popular tool to study IDPs[28-30]. Furthermore unlike many other biophysical techniques, MS is well suited to the analysis of complex mixtures.

The ability of mass spectrometry to distinguish changes in the tertiary fold of a protein based on the number of charges a protein ion can accept during the electrospray process has been long established, with more unfolded species generally presenting to the gas-phase with a greater numbers of charges[31,32]. Hence it is inherently well suited to the study of IDPs which are often conformationally heterogeneous. A study by Frimpong *et al*[33] into the IDP α-synuclein showed that by charge deconvolution of the mass spectra, different conformational families could be observed for this disordered system. Furthermore, by altering solution conditions such as pH and solvent composition the relative intensities of these overlapping families could be altered. This demonstrates that different conformational states of a heterogeneous IDP such as α-synuclein could be preserved and studied in the gas-phase. Studies by Iakoucheva *et al*[34] showed that using proteolytic digest methods coupled with MS, it is actually possible to predict which regions of a protein are likely to be ordered or disordered, with disordered cleavage sites being favourably and commonly cleaved, whereas cleavage in sites of structural order in a protein are less favourable. The results of their study compared favourably to areas predicted by the neural network program predictor of natural disordered regions to be disordered. Furthermore, the combination of ion mobility with mass spectrometry has proved incredibly useful as a tool to study the conformations of intrinsically disordered proteins[35-37]. MS-based studies have also proved

extremely valuable in studying binding of IDPs to ligands or other binding partners, providing information on stoichiometry of binding and conformations adopted by bound species[38-42]. These studies have suggested that gas-phase, mass spectrometry-based techniques have the potential to be a useful tool for studying disordered, dynamic systems.

This proof-of-concept study has focused on the leucine zipper regions of c-MYC (c-MYC-ZipΔDT) and MAX (MAX-Zip), and a longer c-MYC peptide containing the full previously identified drug binding region (c-MYC-Zip), Figure 3.1C. CD is employed in combination with MS and DT IM-MS to study their interactions, in the absence and presence of the inhibitor 10058-F4. The aim of the study was to determine if this is a feasible method for the study of interactions of c-MYC with future ligands, designed to be more potent, and with further development aimed at studying the full length proteins.

3.2 Experimental

3.2.1 Peptide and ligand samples

Peptides of the leucine zipper regions of c-MYC (with and without the drug target region) and MAX where synthesised by the MacMillan research group (University College London, London, UK). The sequences of the peptides synthesised are shown in Figure 3.1C. Details of the synthetic procedure implemented by the MacMillan research group has been published elsewhere[43], brief details are given in Appendix 2 section A2.1. Peptides were supplied as lyophilised samples and stored at -20 °C until use.

Stock solutions of peptides were typically prepared at 350 or 400 μM concentration in 20 mM ammonium acetate (AmAc), pH 6.8. Ammonium acetate salt was purchased from Fischer scientific (Loughborough, UK) and prepared in high purity water, in the resistivity range of 18.0-18.2 MΩ•cm, obtained from an Arium 611 water purification system (Satorius, Göttingen, Germany). Peptide concentrations were checked using the Pierce BCA concentration assay (Thermo Fischer Scientific, Wilmington, DE, USA) following the manufacturer's instructions but using volumes 10 times lower than stated in the manual, afforded through the use of low volume UV cuvettes. Peptides were stored at 4 °C and diluted to the desired concentration before use.

The ligand 10058-F4, used in c-MYC:MAX studies was purchased from Sigma-Aldrich Company Ltd (Dorset, UK), 1 and 0.1 mg/mL stock solutions were prepared in acetonitrile (Sigma Aldrich company Ltd, Dorset, UK) and stored in the fridge at 4 °C until use.

3.2.2 Sample preparation for CD studies

For circular dichrosim spectroscopy (CD) studies single peptide solutions were prepared at 25 μM in 10 mM AmAc buffer, pH 6.8. The mixed peptide solutions were prepared at a 1:1 ratio with a total protein concentration of 25 μM in 10 mM AmAc buffer, and incubated for 1 hour at 37 °C before analysis. Mixed peptide plus 10058-F4 ligand samples were prepared at a 1:1:1 ratio in which each component was present at 12.5 μM, and incubated for one hour at 37 °C before analysis. In order to study the effect of pH on the observed CD spectra of c-MYC-Zip plus MAX-Zip, with and without the ligand (10058-F4), samples were prepared at 1:1 or 1:1:1 ratios in 10 mM AmAc, buffer pH was adjusted to 9.8 or 2.8 through the drop-wise addition of either sodium hydroxide (NaOH) solution (Sigma Aldrich Company Ltd, Dorset, UK) or formic acid (VWR International Ltd, Leicestershire, UK).

3.2.3 Sample preparation for MS studies

MS studies were performed on c-MYC-Zip plus MAX-Zip and c-MYC-ZipΔDT plus MAX-Zip, prepared in 20 mM AmAc with 20 % acetonitrile at a 1:1 ratio, with each peptide present at a concentration of 125 μM. 10058-F4 was prepared in acetonitrile hence addition of acetonitrile to the mixed peptide solutions enables any effect of the solvent to be distinguished from ligand affects. Mixed peptide plus 10058-F4 samples were prepared at equimolar concentrations in 20 mM AmAc, with all components present at 125 μM. All samples were incubated at 37 °C in a digital dry bath (Jencons-PLS, East Grinstead, UK) and analysed after 15 minutes, one or three hours. For c-MYC-Zip plus MAX-Zip plus 10058-F4 two additional incubation times were studied, after six and 24 hours at 37 °C. For c-MYC-Zip plus MAX-Zip plus 10058-F4 the effect of altering buffer strength over the range 10-100 mM AmAc was determined. In addition at 20 mM AmAc the effect of altering the pH over the range 9.8 to 2.8 was studied. Finally, this mixed peptide solution was prepared at 1:1:5 (c-MYC-Zip:MAX-Zip:10058-F4) ratio in 20 mM AmAc pH 6.8. All MS studies presented in this chapter were performed on a Q-TOF II mass spectrometer (Waters. Manchester, UK) as described in Chapter 2 section 2.2.

3.2.4 Sample preparation for DT IM-MS studies

For determination of collision cross sections (CCS), single peptide solutions were prepared at a concentration of 125 μM. Mixed peptide solutions were prepared at 1:1 ratios with each peptide present at 125 μM. In the presence of ligand, samples were prepared at equimolar concentrations with each component at a concentration of 125 μM. In all cases samples were prepared in 20 mM AmAc, for mixed peptides in the absence of ligand 20 % acetonitrile was

added. Samples were incubated at 37 °C before analysis, in the absence of ligand all samples were incubated for 1 hour and in the presence of the ligand samples were incubated for 3 hours. Furthermore CCS of c-MYC-Zip:MAX-Zip in the presence of 10058-F4 was studied at an additional time point, after 1 hour incubation. All DT IM-MS measurements were performed on an in-house modified Q-TOF (Micromass Uk Ltd.), adapted in order to carry out separations based on an ion's mobility, and to enable the temperature-dependent CCS to be determined, as described in Chapter 2 section 2.3.

3.2.5 Molecular dynamics simulations

All molecular dynamics simulations and theoretical calculations discussed in this chapter were performed by Dr Massimilano Porrini. Details of the procedure can be found elsewhere[43] and are given in brief in Appendix 2 section A2.2. Molecular dynamics simulations were run to enable the theoretical CCS to be determined for the c-MYC-ZipΔDT:MAX-Zip hetero-dimeric leucine zipper complex (PDB code 1A93). The disordered monomeric and dimeric species CCS were calculated following simulated annealing of the monomers or heterodimer.

3.3 Results and discussion

3.3.1 Solution-phase secondary structure analysis: circular dichroism spectroscopy

CD spectroscopy was utilised in order to study the secondary structure content of the individual peptides and peptide mixes before and after ligand addition (Figure 3.2). For each individual peptide solutions the observed CD spectrum is consistent with that of a system containing a proportion of α-helical and coiled regions (Figure 3.2A and Table 3.1). This is in contrast to the reports of Hammoudeh *et al*[17] on similar peptide regions of c-MYC in which they only saw an α-helical signature upon addition of c-MYCs binding partner MAX; however, slightly differing peptide regions of the c-MYC:MAX interaction domain were studied here[17,22]. It is not possible, however, to determine from CD if this helicity is due to monomers or homodimers. Homodimers of c-MYC are thought to be rare, but they have been previously reported[44]. Previous studies have reported the use of CD to study homodimers of c-MYC[24,45], these studies have focused on peptides of the leucine zipper region but have often covalently attached the monomeric chains to enable the dimeric structure to be studied in this way[24]. The studies highlighted that the homodimers adopt

α-helical, sometimes coiled-coil, structures and that even monomeric c-MYC has some helicity. MAX homodimers are, however, known to be more favourable and functional, being capable of binding DNA. Furthermore their structure has been solved using high resolution techniques and shown to be consistent with that of the helical b-HLH-Zip structure, hence homodimers could account for the helical signature observed here for MAX[45,46]. It is therefore plausible that the helical signature seen here for all individual peptides could be due to monomeric species or dimeric species or a combination of both; however, from CD alone it is impossible to determine which species the solution spectra are dominated by.

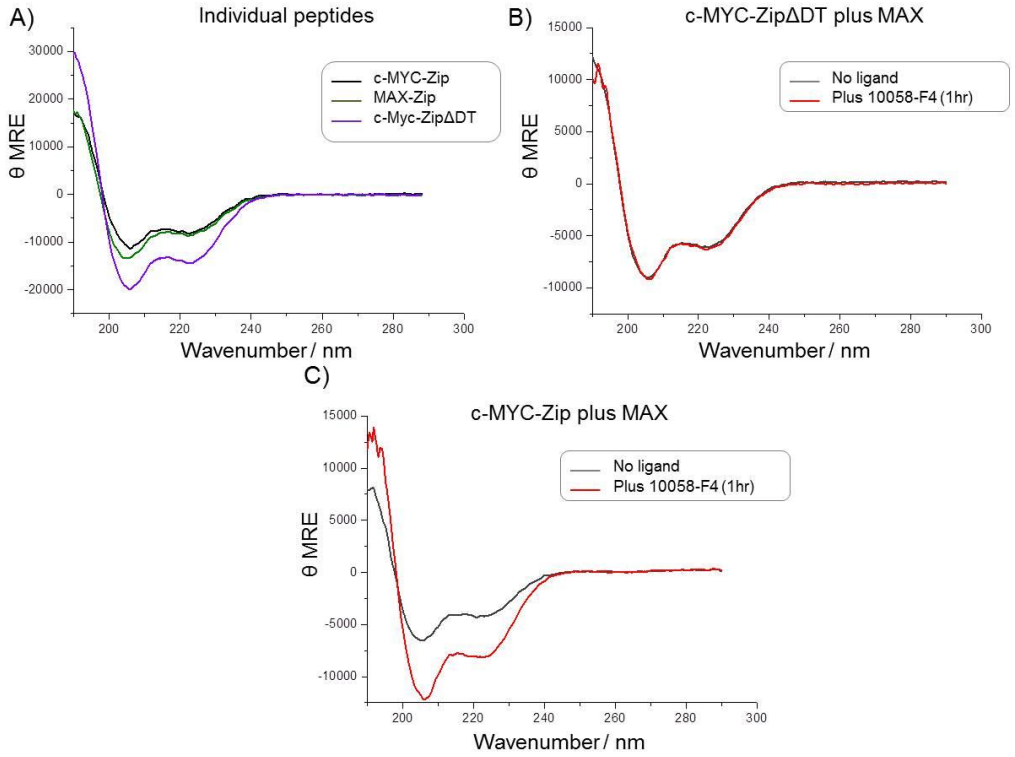

Figure 3.2: CD spectra obtained for A) individual peptide solutions of c-MYC-zip (black), MAX-Zip (green) and c-MYC-ZipΔDT (purple). B) c-MYC-ZipΔDT plus MAX in the absence of 10058-F4 (grey) and after one hour incubation (red) with 10058-F4. C) c-MYC-Zip plus MAX in the absence of 10058-F4 (grey) and after one hour incubation (red) with 10058-F4. MRE was determined in individual peptide mixes based on peptides being monomeric and in peptide mix solutions assuming all peptides were in the heterodimeric form.

c-MYC-ZipΔDT plus MAX (Figure 3.2B and Table 3.1) and c-MYC-Zip plus MAX-Zip (Figure 3.2C and Table 3.1) mixtures provide similar α-helical signatures, as expected for the heterodimeric complexes. Interestingly, this helical signature is calculated to be lower than that observed for the single peptide solutions and for c-MYC-Zip:MAX is seen to increase in the presence of the ligand 10058-F4. For both mixes the percentage of disordered content increases also, suggesting either that the complexes formed by these peptides contain

a certain amount of disorder or that the unbound peptides are substantially disordered in this mixed peptide solution. It is important to note that in the case of the mixed peptide solutions, the mean residue ellipticity (MRE) was calculated based on all peptides in the solution forming a heterodimeric complex; however, it is likely that the process is more complicated than this with some proportion of unbound monomer or perhaps even homodimers being present in the solution.

Solution	***Helical content / %***	***Sheet content / %***	***Coil content / %***	***Disordered content / %***
c-MYC-Zip	20	29	12	37
c-MYC-ZipΔDT	44	11	15	3
MAX-Zip	22	27	13	37
c-Myc-ZipΔDT plus Max	8	37	11	43
c-Myc-ZipΔDT plus Max plus 10058-F4 (1 hr)	9	36	11	43
c-Myc-Zip plus Max	5	41	12	41
c-Myc-Zip plus Max plus 10058-F4 (1 hr)	20	29	14	36

Table 3.1: The secondary structure content as determined from the CD spectra shown in Figure 3.1, calculated from DICHROWEB server using the CDSSTR algorithm[47-49]. A representative example of the quality of fitting is given in Appendix 2 Figure A2.1. In all cases data was input as normalised to MRE and single peptide solutions were considered to be monomeric species and mixed peptide solutions were considered as heterodimers.

Further CD experiments were performed on the c-MYC-Zip:MAX-Zip peptide mixture at pH 9.8 and pH 2, in the absence and presence of the ligand 10058-F4, (Appendix 2 Figure A2.2). At both pHs in the absence and presence of the ligand, the helical content is found to be higher than at pH 6.8. This is unsurprising as the sodium hydroxide used for pH adjustment (pH 9.8) can help stabilise salt bridges in a helical form, furthermore dimerization is thought to be promoted under acidic conditions[50].

Interestingly, although in all cases the CD spectra visually appear to be mostly helical, secondary structural analysis using the CDSSTR algorithm in DICHROWEB reveals a substantial proportion of disorder and sheet content in these peptides, both in the individual solutions and the mixed peptide solutions. It is clear from the CD experiments performed here that CD provides ambiguous results for a complex multicomponent, conformationally dynamic system and in order to probe the species present in such samples a technique which can distinguish between species of different oligomeric orders and different conformations is needed. In order to further study this biologically important system the peptides were then

studied by MS and DT IM-MS, which are intrinsically well suited to study conformationally dynamic systems.

3.3.2 Investigating ligand binding: a mass spectrometry approach

The mass spectra for the individual peptides were first acquired and show mostly monomeric species (M) with a relatively narrow charge state distribution, over the range $[M+3H]^{3+}$ to $[M+4H]^{4+}$ for each of the peptides (Figure 3.3).

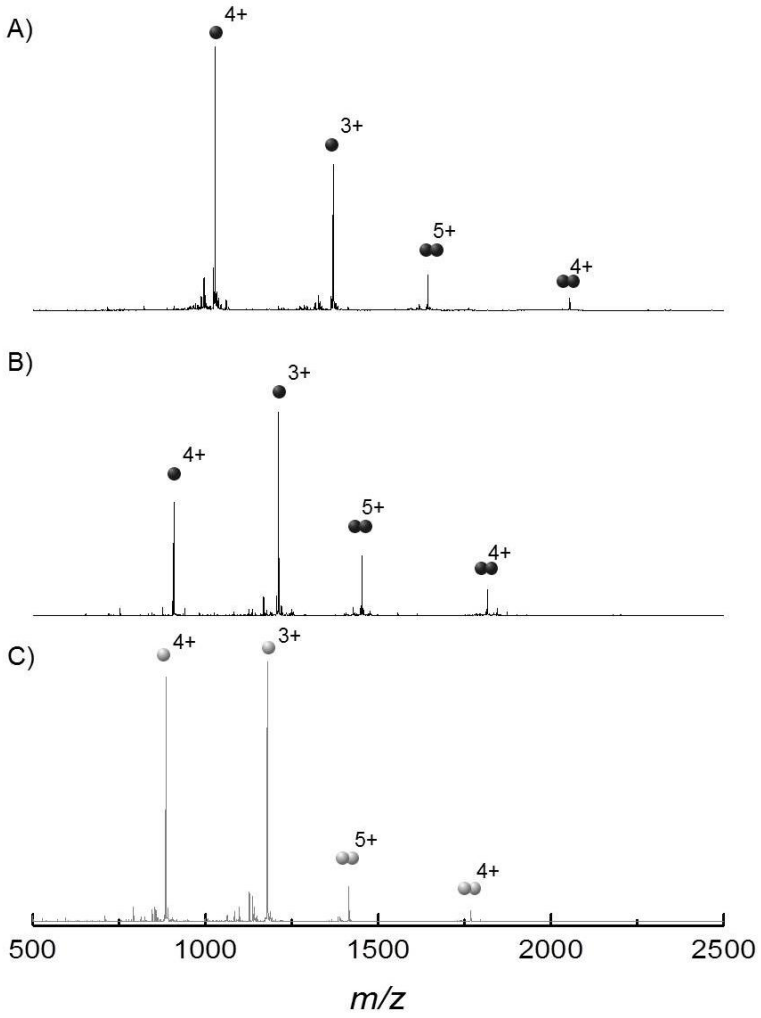

Figure 3.3: Mass spectra obtained from 125 µM peptide solutions in 20 mM AmAc. A) c-MYC-Zip B) c-MYC-ZipΔDT and C) MAX-Zip. All major peptide peaks are labelled with charge states and a sphere represents one monomeric unit.

Theoretical PI values for c-MYC-Zip, c-MYC-ZipΔDT and MAX are as follows; 6.79, 6.76 and 8.36, and correspond to net charges of 0, 0, and +2 respectively at pH 7. Here experiments were performed at pH 6.8 and it is observed that the ions typically populate lower charge states than the total number of protonatable sites, suggesting all potential sites are not accessible for protonation which could be due to sites not being solvent accessible

during ionisation. Furthermore it is clear that all peptides studied here are capable of homodimerisation, with homodimers (D) of c-MYC-Zip, c-MYC-ZipΔDT and MAX-Zip, being observed in each case as the $[D+4H]^{4+}$ and $[D+5H]^{5+}$ species. Homodimerisation accounts for approximately 10, 24 and 8 % of the total intensity of the peptide species observed by MS for c-MYC, c-MYC-ZipΔDT and MAX-Zip respectively. In all cases the $[D+4H]^{4+}$ could also be due to the monomeric species $[M+2H]^{2+}$, as they have the same mass-to-charge ratio, however, the isotope distribution for these peaks shows, for all peptides, they are primarily due to the dimer and hence have been identified as such. Additionally a number of lower intensity peaks are present in all spectra due to contaminants carried over from the synthetic procedure.

The peptide mixtures c-MYC-Zip plus MAX-Zip and c-MYC-ZipΔDT plus MAX-Zip were examined to determine if the heterodimeric complex could be formed, preserved and transmitted into the gas-phase. Samples were mixed and incubated for 15 minutes, one or three hours before analysis. For both peptides, at each incubation time, the spectra obtained showed that the most intense species were monomers over the charge state range $[M+3H]^{3+}$ to $[M+5H]^{5+}$, Figure 3.4 shows spectra obtained after three hours incubation. The increase in the highest observed charge state from $[M+4H]^{4+}$ to $[M+5H]^{5+}$, for the monomeric species in the mixed peptide solutions is attributed to the addition of 20% acetonitrile to the solution in these studies. Acetonitrile was used here to replicate the solvent conditions employed for ligand binding studies where 10058-F4 was prepared in acetonitrile and this increase in charge state distribution is observed for both the c-MYC-Zip and c-MYC-ZipΔDT individual peptide solutions plus 10058-F4 (Appendix 2 Figure A2.3), attributed to the acetonitrile used to prepare the ligand.

The c-MYC-Zip:MAX-Zip and c-MYC-ZipΔDT:MAX-Zip heterodimeric complexes (C), however, were also observed, as $[C+4H]^{4+}$and $[C+5H]^{5+}$. Upon the addition of stoichiometric amounts of 10058-F4 (1:1:1), no ligand:peptide or ligand:complex species were observed for either c-MYC peptide. Furthermore, the single c-MYC peptide solutions were also studied in the presence of the ligand 10058-F4 at a 1:1 ratio and again no ligand:peptide complexes were observed for either c-MYC peptide, with (c-MYC-Zip) or without (c-MYC-ZipΔDT) the drug target binding region (Appendix 2 Figure A2.3). A dissociation constant (K_d) of 5.3 μM[17] has been reported for the 10058-F4 ligand to c-MYC and additional work highlighted that truncated peptide regions containing the DT region bind with a similar affinity[22]. These reported K_d values should be sufficient to preserve *via* n-ESI[51].

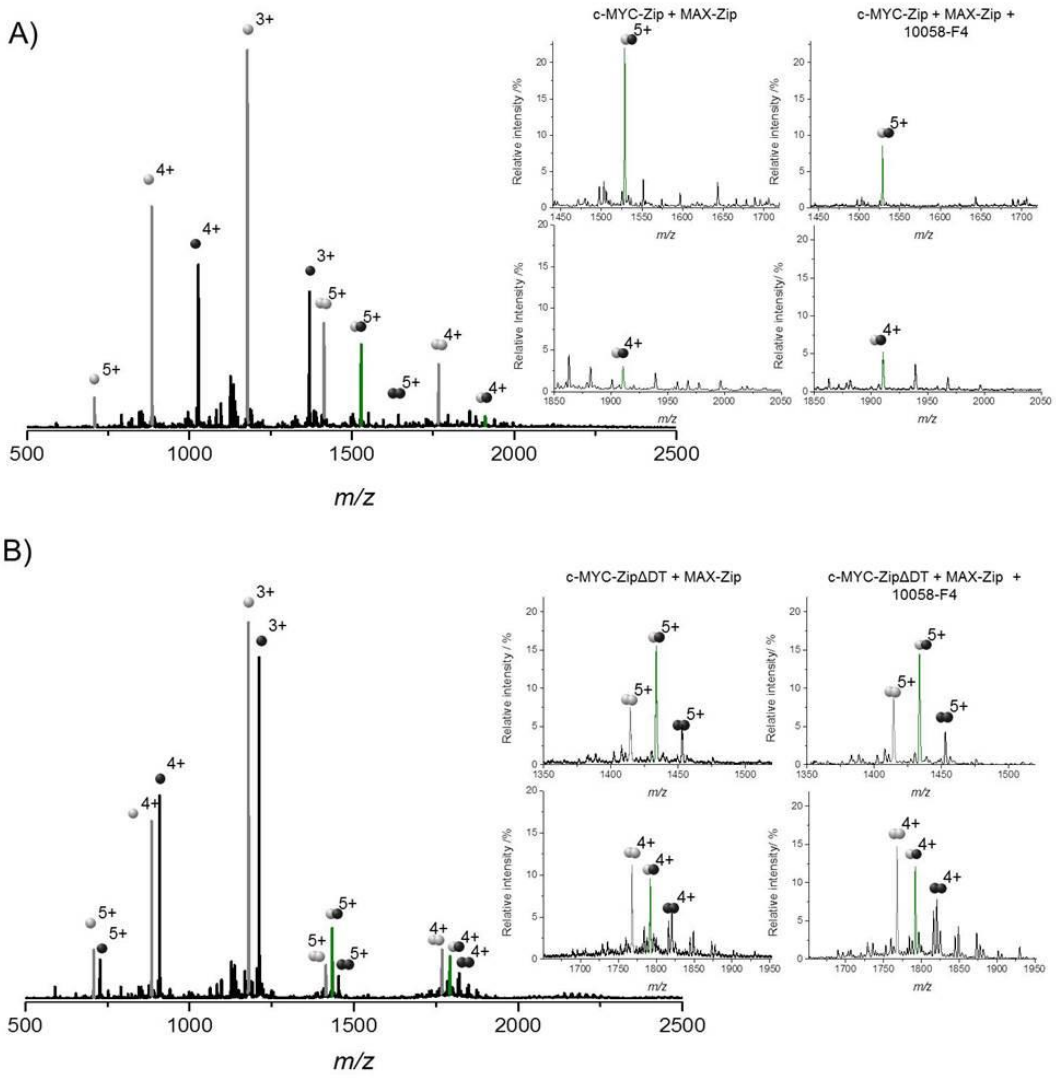

Figure 3.4: Mass spectra recorded after 3 hours incubation at 37 °C, peptide mixtures were prepared at a 1:1 ratio at 125 µM of A) c-MYC-Zip plus MAX-Zip. Insert: magnified region of the spectra showing the c-MYC:MAX-Zip complexes in the absence and presence of 10058-F4 B) c-MYC-ZipΔDT plus MAX-Zip Insert: magnified region of the spectra showing the c-MYC-ZipΔDT:MAX-Zip complex in the absence and presence of 10058-F4. In both cases peaks due to c-MYC peptides are in black and peaks due to MAX-Zip are in grey, green peaks are heterodimeric species.

In order to try to observe the peptide:ligand or complex:ligand species a number of different solution and desolvation conditions were employed for MAX-Zip plus c-MYC-Zip, the peptide which contains the previously identified drug target binding region. Firstly, samples were incubated at room temperature for 15 minutes, one hour or three hours but again no ligand complexes were observed following these experiments (Appendix 2 Figure A2.4).

The effect of varying the solution conditions was then probed by changing both the concentration of the ammonium acetate buffer, over the range of 10-100 mM, and the pH of the solution, using 20 mM ammonium acetate over the pH range 9.8 to 2.8 (Appendix 2

Figures A2.5 and A2.6 respectively). Under all solution conditions studied here no 10058-F4:c-MYC-Zip complex or 10058-F4:c-MYC-Zip:MAX-Zip complexes were observed. The addition of the ligand itself does not cause any significant shift in the pH of the sample, however, it is less soluble at low pH, crashing out of solution below pH 2.8 at concentrations of 125 μM.

Next a number of desolvation and ionisation conditions were probed including the effect of capillary voltage, the potential applied in order to ionise and desolvate the sample. Typically a capillary voltage of 1.6 kV was used in the studies reported here for c-MYC-Zip plus MAX-Zip, however, this applied voltage was varied over the range 1.3-2.0 kV to determine if the applied voltage could influence the ability to preserve and transmit any ligand-bound complexes (Appendix 2 Figure A2.7). At all applied capillary voltages no ligand-bound species were observed, however. Furthermore, the effect of cone voltage was investigated over the range 10-40 V, but even with the gentlest conditions no ligand complex could be observed (Appendix 2 Figure A2.8).

The final technique employed in an attempt to preserve any ligand-bound complexes was solvent vapour exposure, a technique which has previously proved successful in stabilising weak gas-phase protein:ligand interactions[52]. Solvent vapour exposure involves inserting a small open container of solvent into the source region of the mass spectrometer, in this case solvent was held in the lid of a falcon tube inserted below the source cone. The falcon tube lid was filled with acetonitrile but again no peptide:ligand complexes were observed (Appendix 2 Figure A2.9).

The fact that no c-MYC-Zip:10058-F4 or c-MYC-Zip:MAX-Zip:10058-F4 complexes were observed under any of the numerous conditions employed here is significant. The ligand is thought to bind c-MYC through hydrophobic interactions between the aromatic ring of 10058-F4 and hydrophobic residues in the drug target (DT) region of c-MYC-Zip[22] and previous work has shown the preservation of hydrophobic protein:ligand interactions is indeed possible in the gas-phase[53,54]. Hence it can be inferred that the K_d of the ligand may be greater than quoted in earlier studies, and any ligand binding in solution is too weak to be preserved during desolvation and ionisation even under very gentle source conditions, a finding which is consistent with recent metadynamics simulations which suggested the binding of 10058-F4 to a c-MYC peptide of the Zip region is driven by weak, non-specific interactions with hydrophobic patches[55].

For both c-MYC peptides a notable decrease in the intensity of the $[C+5H]^{5+}$ complex species was, however, observed in the presence of the ligand, with the largest decrease in complex signal observed for the c-MYC-Zip:MAX-Zip complex being after three hours incubation at 37 °C, (Figure 3.4 insert and Table 3.2). An increase in the $[C+4H]^{4+}$ species, however, is also observed (Figure 3.4 insert and Table 3.2). Similar observations are made for both c-MYC-Zip:MAX-Zip and c-MYC-ZipΔDT:MAX-Zip, despite the latter not containing the full drug target region, again suggesting interactions between 10058-F4 and c-MYC are not as specific as initially described. From MS experiments alone it appears that 10058-F4 does not significantly perturb the total equilibrium between free peptide and heterodimeric complex but it does affect the charge state distribution of the complex. This shift in charge state could be indicative of a shift in the conformational equilibrium of this peptide complex, which can be probed directly utilising DT IM-MS.

	c-MYC-Zip:MAX				***c-MYC-ZipΔDT:MAX***			
Incubation	No 10058-F4		Plus 10058-F4		No 10058-F4		Plus 10058-F4	
time	5+	4+	5+	4+	5+	4+	5+	4+
15 minutes	11.4	3.1	10.1	5.5	23.2	6.4	8.3	3.9
1 hour	16.0	2.4	8.9	6.1	18.0	7.8	10.6	12.5
3 hours	18.7	2.7	5.21	5.2	15.7	9.3	12.1	9.9

Table 3.2: The relative intensities of the c-Myc-Zip:MAX-Zip and c-Myc-ZipΔDT:MAX-Zip 4+ and 5+ complexes, after incubation at 37 °C for 15 minutes, 1 hour and 3 hours, in the absence and presence of the ligand 10058-F4.

3.3.3 Studying conformational change utilising ion mobility-mass spectrometry

The conformations adopted by the individual peptides, homo- and heterodimers were studied using DT IM-MS both in the absence and presence of 10058-F4. Table 3.3 contains the calculated experimental CCS determined in the absence of the ligand. Experimental CCS were determined for all observed species and compared to theoretical CCS obtained either from the NMR structure (PDB 1A93)[20] or from a simulated annealing (SA) procedure. The theoretical values obtained from both procedures are also shown in Table 3.3.

From DT IM-MS experiments, for all species studied, an increase in CCS with charge is observed, as expected, as more unfolded species will be able to accept larger numbers of charges. Furthermore, for small peptide systems like those studied here an increase in charge is likely to increase the CCS due to Columbic repulsion of proximal charges, which is more probable to have a significant effect on small peptide systems where charges are likely to be inherently close to each other.

Species		*Experimental CCS, Å²*	*Theoretical CCS, Å²*
c-Myc-Zip	M		627*
	$[M+4H]^{4+}$	711 ± 2	
	$[M+3H]^{3+}$	649 ± 3	
	$[M+2H]^{2+}$	479 ± 8	
	$[D+5H]^{5+}$	1131 ± 5	
c-Myc-ZipΔDT	M		592*
	$[M+4H]^{4+}$	629 ± 4	
	$[M+3H]^{3+}$	595 ± 5	
	$[M+2H]^{2+}$	453 ± 11	
	$[D+5H]^{5+}$	941 ± 8	
Max-Zip	$[M+2H]^{2+}$		560*
	$[M+4H]^{4+}$	565 ± 2	
	$[M+3H]^{3+}$	554 ± 6	
	$[D+4H]^{4+}$	857 ± 12	
c-Myc-Zip:MAX-Zip	$[C+2H]^{2+}$		962* 1247‡
	$[C+5H]^{5+}$	982 ± 7	
		1164 ± 9	
	$[C+4H]^{4+}$	923 ± 9	
c-Myc-ZipΔDT:MAX-Zip	$[C+2H]^{2+}$		918* 1162‡
	$[C+5H]^{5+}$	947 ± 3	
		1136 ± 12	
	$[C+4H]^{4+}$	906 ± 3	

Table 3.3: Experimental and theoretical CCS for c-MYC-Zip, c-MYC-ZipΔDT and MAX-Zip monomers (M) and homodimers (D) in addition to the heterodimeric complexes (C); c-MYC-Zip:MAX-Zip and c-MYC-ZipΔDT:MAX-Zip. Experimental CCS listed here were determined from single peptide, and mixed peptide solutions in the absence of the ligand. Theoretical CCSs were determined either from simulated annealing (*) or from NMR coordinates (‡).

For the individual monomeric peptides it is observed that c-MYC-Zip adopts the largest conformations with CCS of ~479-711 $Å^2$ over the charge state range $[M+2H]^{2+}$ to $[M+4H]^{4+}$. It would be expected that this peptide would adopt the largest CCS out of the three individual peptides as it has the longest chain length, consisting of 33 amino acids whilst c-MYC-ZipΔDT and MAX-Zip are both composed of 29 amino acids. c-MYC-ZipΔDT exists in a similar CCS range over the same charge state range observed for c-MYC-Zip, being only marginally smaller (~453-629 $Å^2$) and is attributed to a similar conformational family to that adopted by c-MYC-Zip. The slightly smaller CCS observed are ascribed to the shorter chain length for this peptide. The CCS of MAX were determined only for $[M+3H]^{3+}$ and $[M+4H]^{4+}$, and are slightly more compact (~554-565 $Å^2$) than those observed for the c-MYC peptides at these charge states, demonstrating that monomeric MAX-Zip is capable

of existing in more compact forms. The conformations adopted by these monomeric peptides in the mixed peptide samples in the presence of 10058-F4, was also studied and CCS of all monomeric species remain the same within experimental error (Appendix 2 Table A2.1).

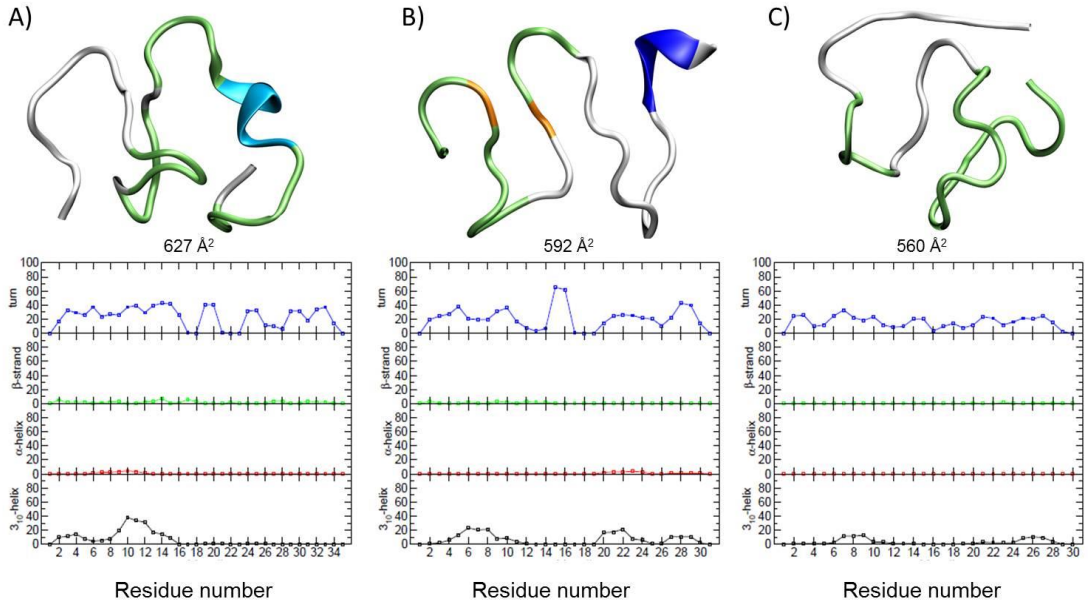

Figure 3.5: A) c-MYC-Zip B) c-MYC-ZipΔDT and C) MAX-Zip. Top: Theoretically predicted structures obtained through SA. In the SA procedure for each stored minimised structure the rotationally averaged CCS was determined and the representative structures were chosen selecting the lowest energy species among those having a CCS "close" to the peak position of the distributions. The CCS of each representative structure is shown below the structure. The colours refer to the secondary structure as follows; green, white, orange, blue and cyan refer to turn, coil, isolated bridge, 3_{10}-helix and α-helix respectively. Bottom: Secondary structure contents in percentage determined considering a window of 10.0 $Å^2$ centred at the CCS distributions peak position, ensembles of structures (~100 structures), using the DSSP algorithm.

Experimental CCS obtained for the monomeric peptides were compared to theoretical CCS obtained from a SA molecular modelling procedure. From Table 3.3 it is clear that the experimentally obtained conformations adopted by these peptides are very similar to those predicted from the SA procedure, representative structures obtained are shown in Figure 3.5 and visually appear very disordered in nature. It is possible, however, to determine the secondary structure contents of these predicted conformations *via* the DSSP algorithm implemented in Amber10[56]. The DSSP algorithm enables the secondary structure of amino acids in a peptide or protein to be assigned based on the atomic-resolution coordinates of the species. The resulting trends for ~100 structures in a 10 $Å^2$ window centred at the peak of the CCS distribution as determined by the SA procedure are shown for the three peptides in Figure 3.5. These monomeric peptides are primarily disordered as would be expected for these species, which have been shown to undergo a disorder-to-order transition upon binding. Interestingly though for all individual peptides the DSSP analysis shows these

species contain residues implicated in the formation of 3_{10}-helices, which could account towards the helical signature observed in CD for all individual peptide solutions studied.

Considering next the heterodimers, Figure 3.6 presents the CCS distributions obtained for $[C+5H]^{5+}$ of both c-MYC-Zip:MAX-Zip (Figure 3.6A) and c-MYC-ZipΔDT:MAX-Zip (Figure 3.6B), in the absence (top) and presence (bottom) of the ligand. For both c-MYC peptides similar observations are made, with striking differences detected by DT IM-MS in the absence and presence of the ligand. Considering first the heterodimeric complexes ($[C+5H]^{5+}$) in the absence of ligand, two conformations are observed (Table 3.1 and Figure 3.6 top), with average CCS of 982 and 1164 $Å^2$ for c-MYC-Zip:MAX-Zip and 906 and 1136 $Å^2$ for c-MYC-ZipΔDT:MAX-Zip. Comparison of these experimental CCS with theoretical calculations can provide candidate conformations. Considering the more extended conformation of both heterodimeric complexes, close agreement to the theoretical CCS calculated for the leucine zipper dimer derived from the NMR co-ordinates is observed (represented by black dotted lines and corresponding structures in Figure 3.6). For c-MYC-ZipΔDT:MAX the theoretical CCS is calculated as 1162 $Å^2$, highlighting that this extended conformation is likely to be the leucine zipper type conformation.

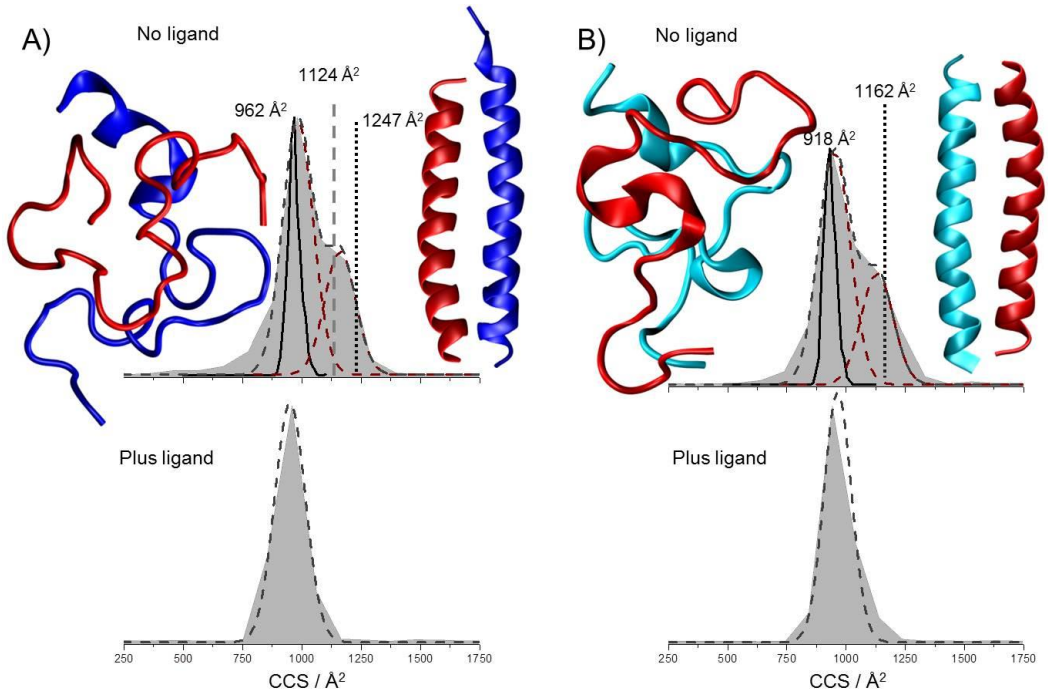

Figure 3.6: A) c-MYC-zip:MAX-Zip B) c-MYC-ZipΔDT:MAX-Zip. CCS distributions derived from arrival time distributions at a drift voltage of 45 V for $[C+5H]^{5+}$. Results in absence of ligand (after 1 hour incubation at 37 °C) and in the presence of the ligand, 10058-F4 (after 3 hours incubation at 37 °C) are shown in the top and bottom panel respectively. Profiles for the predicted conformation from SA (solid black lines) and from NMR structures (vertical black dotted line) are displayed together with experimental fitted curves (dashed lines), CCS obtained from additional SA procedure is shown for c-MYC-Zip:MAX (grey dashed line).

For c-MYC-Zip:MAX the theoretical CCS for the leucine zipper dimer was calculated as 1247 $Å^2$ (represented by black dotted line in Figure 3.6A) however, in order to create the c-MYC-Zip:MAX leucine zipper complex an additional four amino acids (YILS) had to be added to the NMR structure to account for the DT region. Therefore to ensure that this slightly altered structure was still stable in this conformation at 300 K and to allow the structure to relax and equilibrate an additional MD simulation was run (see Appendix 2 section 2.2). The MD simulation consisted of thermalizing and equilibrating the leucine zipper in explicit water, following which a 10 ns simulation was carried out and 100 sample structures were stored (1 every 100 ps) along the obtained trajectory. These structures were then stripped of water and minimised *in vacuo*, the CCS of each resulting structure was calculated and the average was determined to be 1124 $Å^2$ (represented by grey dashed line in Figure 3.6A). More compact than the value obtained prior to relaxation (1247 $Å^2$), however, both are in very close agreement with the experimental value of 1164 $Å^2$. This highlights that this helix:helix interaction, in c-MYC-Zip:MAX, is stable *in vacuo* at the temperature employed during the experimental procedure and is indeed most likely the more extended conformation observed experimentally. This fully helical complex could account for a proportion of the helical signature observed for mixed peptide samples in CD studies.

The earlier arriving conformer observed for both heterodimeric complexes in the absence of the ligand for $[C+5H]^{5+}$ corresponds to a more compact complex (Figure 3.6 top), with a CCS similar to the most visited CCS value of the low energy candidates determined through the SA approach, representative structures are given in Figure 3.6 and CCS distribution obtained are represented by solid black lines. The difference in width observed for the ATDs obtained experimentally as opposed to the distribution obtained through the SA procedure is likely due to both experimental diffusion and conformational heterogeneity, that is to say the structures will exist in a conformational family of similar structures. As with the monomeric structures determined by SA the secondary structure of the heterodimeric structures can be assessed using the DSSP algorithm in Amber10 in this case considering ~160 structures (Figure 3.7). Despite their largely disordered conformation it is clear the heterodimeric complexes predicted through SA contain some regions of 3_{10}-helix and turns, where a turn is a secondary structural element where the polypeptide chain reverses its overall direction and comprises stabilising hydrogen bonds typical of helices.

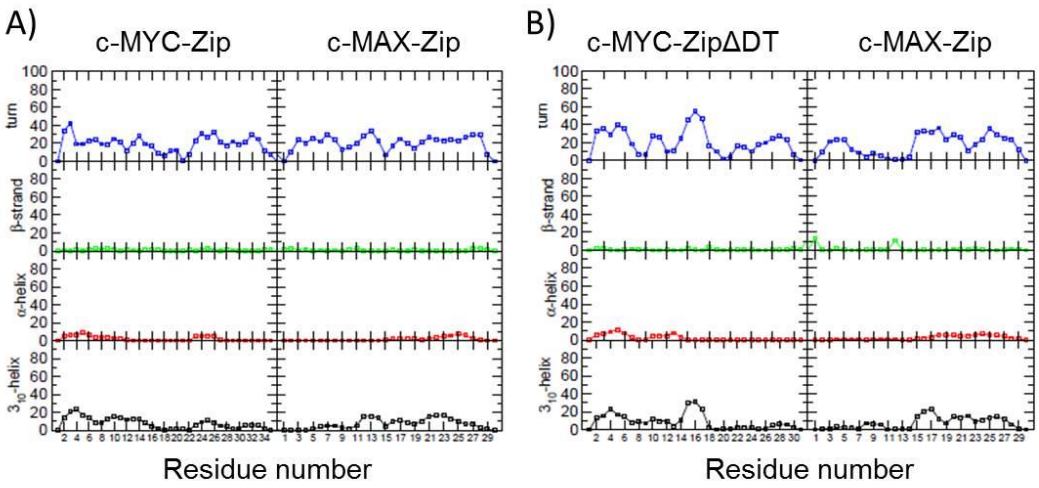

Figure 3.7: A) c-MYC-Zip:MAX-Zip B) c-MYC-ZipΔDT:MAX-Zip. Secondary structure contents in percentage of the disordered dimer complexes predicted from SA, determined considering a window of 10.0 Å^2 centred at the CCS distributions peak position, ensembles of structures (~160 structures) and using the DSSP algorithm. Where, a turn is a secondary structural element where the polypeptide chain reverses its overall direction and comprises stabilising hydrogen bonds typical of helices.

Experimental conditions including pressure in the drift cell, energy with which ions are injected into the cell and time ions spend in the gas-phase are tuned in these DT IM-MS studies, to enable gentle transfer of ions and to cause minimal disruption to solution-phase topologies[57-59]. Therefore, combining experimental results with theoretical CCS determined from SA and from the NMR structure it is surmised that the complex exists in solution as both a structured leucine zipper conformation and a disordered conformation; these disordered conformations could potentially be the encounter complex *en route* to the formation of the highly ordered form. Both conformations can be preserved and transferred into the gas-phase, where DT IM-MS can distinguish between them, highlighting the elegance of this technique to differentiate between conformations in a complex dynamic system. Furthermore, both the fully helical leucine zipper and the disordered complex would contribute to the helical signature seen in the CD spectra for the peptide mixtures adding to the complexity of these spectra.

Upon addition of the ligand 10058-F4 and incubation at 37 °C for three hours the ATDs for $[C+5H]^{5+}$ in both peptide mixtures (Figure 3.6 bottom) appears notably different to those obtained in the absence of the ligand. For both heterodimeric complexes at this charge state in the presence of 10058-F4 only a single conformation is observed, consistent with the compact disordered species observed in the absence of the ligand. The leucine zipper conformation however is no longer observed. This provides persuasive evidence for the inhibition of the helical leucine zipper structure in the presence of 10058-F4, a finding which

could not have been determined by MS or CD alone. The mechanism by which this complex is inhibited is unclear, however, the striking differences in the absence and presence of 10058-F4 suggest that the ligand is interacting either with the c-MYC peptides or the heterodimer in solution presumably stabilising the disordered complex and inhibiting formation of the helical coiled-coiled conformation. These interactions are too weak, or perhaps non-specific, to enable any ligand-bound species to be detected by MS, however, the conformational consequences can be distinguished by DT IM-MS. The fact that similar observations are made for c-MYC-Zip and c-MYC-ZipΔDT peptide, despite the latter not having the DT region, suggests that this ligand 10058-F4, is not as specific as previously reported.

The ATDs shown in the top and bottom panels of Figure 3.6 were, however, obtained after different incubation times, in both cases incubation was at 37 °C. The mixed peptide samples were incubated for one hour before analysis; however, in the presence of the ligand the samples were incubated for three hours prior to analysis, as MS studies showed the greatest decrease in the $[C+5H]^{5+}$ heterodimeric complex at this time point. Therefore in order to confirm that the leucine zipper complex is inhibited due to the presence of the ligand and this observation is not an artefact of increased incubation time, the c-MYC-Zip plus MAX-Zip plus 10058-F4 mixed peptide solution was studied after one hour incubation at 37 °C. In this case the extended species was again absent and only the compact disordered species was observed confirming that that inhibition of this leucine zipper complex is due to the presence of 10058-F4 (Figure 3.8).

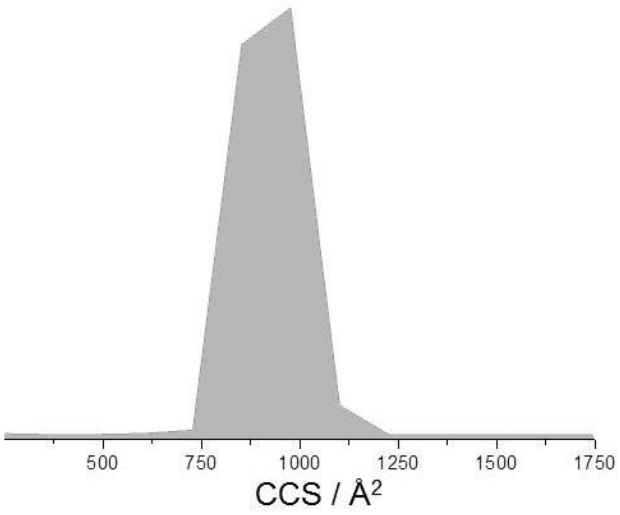

Figure 3.8: CCS distribution derived from arrival time distributions at a drift voltage of 45 V for the c-MYC-ZIP:MAX-Zip complex $[C+5H]^{5+}$ ion after incubation with 10058-F4 at 37 °C for one hour.

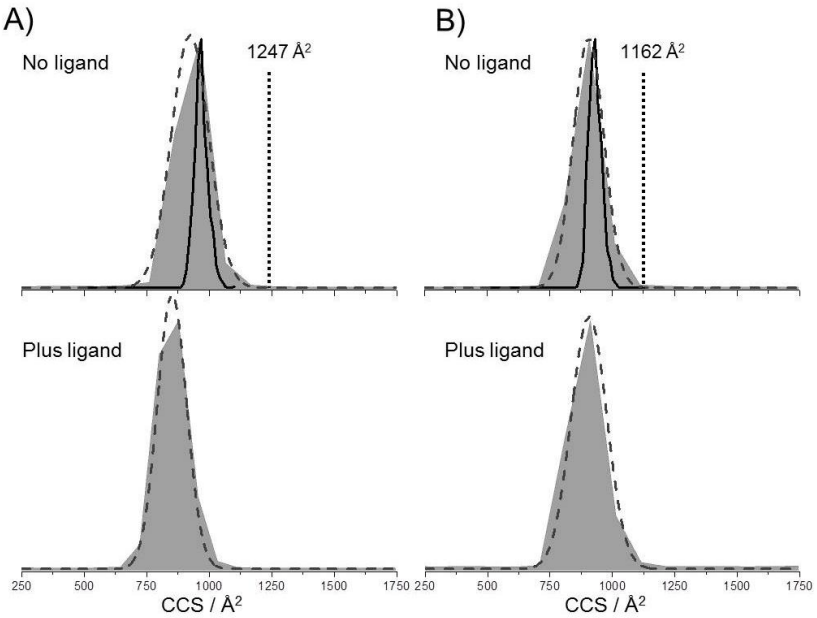

Figure 3.9: A) c-MYC-zip:MAX-Zip B) c-MYC-ZipΔDT:MAX-Zip. CCS distributions derived from arrival time distributions at a drift voltage of 45 V for the complex $[C+4H]^{4+}$ ions. Results in absence of the ligand (after 1 hour incubation at 37 °C) and in the presence of the ligand, 10058-F4 (after 3 hours incubation at 37 °C) are shown in the top and bottom panel respectively. Profiles for the predicted conformation from SA (solid black lines) and from NMR structures (vertical black dotted line) are shown.

The heterodimeric complex is also observed at the 4+ charge state, DT IM-MS experiments show that this ion presents in a single conformational family with a CCS similar to that of the disordered complexes predicted through SA. For both c-MYC peptides the shape of the ATD for this complex does not alter in the presence of ligand (Figure 3.9). The c-MYC-Zip:MAX-Zip $[C+4H]^{4+}$ complex does, however, shift to a slightly lower CCS, this further supports the theory that the 10058-F4 ligand promotes more compact forms of the complex to be favoured in solution, consistent with the findings for $[C+5H]^{5+}$ in which a shift to the more compact, disordered conformation of the complex upon ligand addition is observed. This observation further suggests that 10058-F4 is indeed interacting either with the c-MYC peptide or the heterodimer in solution, causing a conformational change to be observed despite no ligand-bound species being detected by MS. The more compact form may favour a lower charge state in the gas-phase since less protonatable sites will be accessible, and in part this observation explains the relative increase of the $[C+4H]^{4+}$ species in the presence of the ligand. As with c-MYC-Zip:MAX-Zip, the c-MYC-ZipΔDT:MAX $[C+4H]^{4+}$ complex exists as a single compact conformation (Figure 3.9). In this case, however, the CCS of the heterodimer is not influenced upon the addition of 10058-F4, with no compaction observed for the complex in this case, perhaps due to the already more compact nature of this species in the absence of the ligand, suggesting that it is already in its

most compressed form or perhaps that it is a more stable species and therefore this disordered complex experiences less stabilisation upon ligand addition and hence no compaction is observed.

The consistent differences in the conformations observed for the complex ions at the two charge states observed is interesting to consider, with the complex being observed as a single conformational family for $[C+4H]^{4+}$ and as two distinctive conformational families for $[C+5H]^{5+}$. Both charge states exist in a compact conformation which compares extremely well with the disordered complex assigned through the SA simulations. The $[C+5H]^{5+}$ species, however, also exists in a more extended conformation, similar to that of a coiled-coiled leucine zipper complex. The fact that this leucine zipper like conformation is only observed at the higher charge state is unsurprising as helical structures with their side chains sticking out would be more likely to exist in higher charge states[60]. The inherent flexibility of the disordered complex would allow it to exist in a wider range of charge states and also, due to self-solvation of the peptides, potentially lower charge states. Furthermore, visual examination of the disordered complex highlights that the chargeable residues reside mostly in the core of the complex and hence this conformation is likely to present at lower charge states (Figure 3.10). In contrast, the leucine zipper is unable to undergo the same self-solvation of charges and is therefore less likely to be observed at lower charge states.

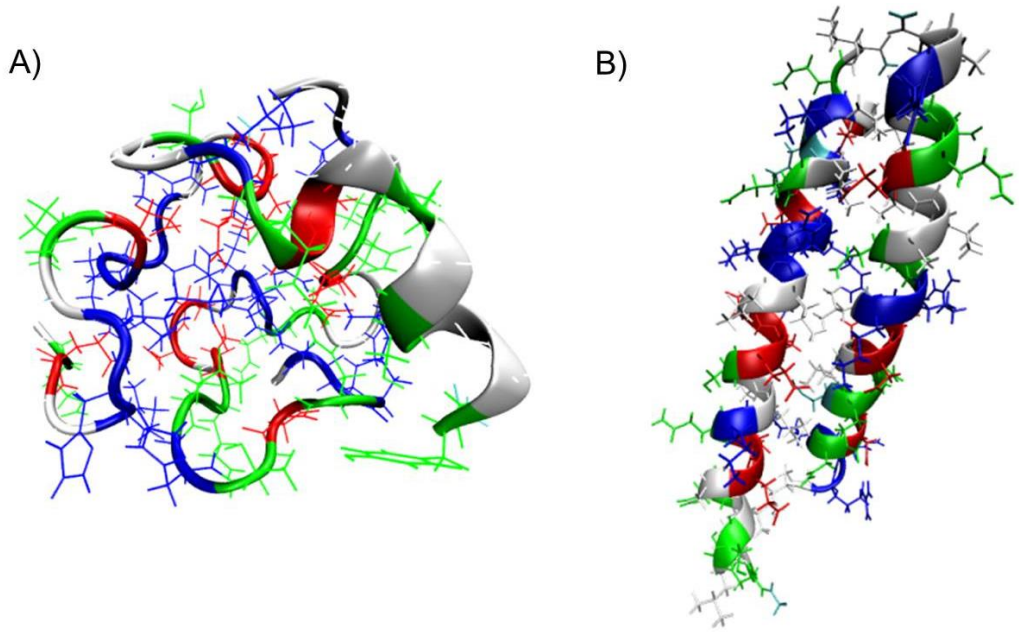

Figure 3.10: c-MYC-Zip:MAX-zip complex structures from A) simulated annealing B) based on NMR structure (PDB 1A93). Colour coded by residue type where non-polar, basic, acidic and polar residues are shown in white, blue, red and green respectively.

The CCS of the homodimers of c-MYC-Zip, c-MYC-ZipΔDT and MAX-Zip were also determined (Table 3.3) and follow the same trend as the monomeric peptides with the MAX-Zip homodimer being the most compact (857 $Å^2$) and the c-MYC-Zip homodimer being the most extended (1131 $Å^2$). The c-MYC-Zip and MAX-Zip homodimers are similar in CCS whilst the c-MYC-Zip homodimer is significantly larger suggesting that the conformation adopted by the homodimer of c-MYC-Zip is significantly different to those adopted by the other homodimers. Comparing the experimental CCS obtained from the homodimers to the theoretical CCS obtained for the heterodimers, it appears that c-MYC-ZipΔDT and MAX both most likely exist in a disordered conformation whilst the c-MYC-Zip homodimer appears to be in the leucine zipper conformation. It would be expected of all the homodimers that MAX would be the one most likely to exist in a leucine zipper conformation due to its propensity to form homodimers[45,46]. However, perhaps the longer chain length of c-MYC-Zip allows this complex to form as a homodimer and it is limited by chain length for other homodimers, an effect which is overcome with the specific binding of the heterodimeric complex, as observed for the $[C+5H]^{5+}$ species. Alternatively, it is possible that the pH of the solutions (pH 6.8) is more conducive for the disordered complex to form for MAX-Zip homodimers and alteration of the pH or addition of limited amounts of salt to the solutions may result in MAX-Zip leucine zipper homodimers being observed by DT IM-MS. In all cases the conformations adopted by the homodimers are not observed to change upon the addition of 10058-F4 (Appendix 2 Table A2.1) suggesting this ligand selectively inhibits the c-MYC:MAX leucine zipper dimers whilst having no effect on the homodimer formation over the time scale considered here, supporting the specificity of this ligand to the disordered heterodimer.

3.4 Conclusions

The studies presented here focused on the leucine zipper regions of c-MYC and MAX, their binding and inhibition by the ligand 10058-F4. Two different length c-MYC peptides were synthesised for this study, both containing the leucine zipper region but the peptide c-MYC-Zip also contained the previously identified 10058-F4 binding site. The studies presented here, however, show no significant differences in complex formation, inhibition or binding of 10058-F4, between the c-MYC-Zip and c-MYC-ZipΔDT peptides despite the lack of the full proposed binding site in the latter. The MS and DT IM-MS results here therefore imply that the ligand is not as specific as previously reported and, based on the fact that no

ligand-bound species are observed, that the ligand has a K_d higher than that previously reported. Supported by a recent study utilising metadynamics simulations on a 10 amino acid stretch of c-MYC containing the identified drug target region (c-$MYC_{402-412}$) which suggested that the ligand binding to c-MYC is driven primarily by weak, non-specific interactions with hydrophobic patches[55]. If these binding interactions are indeed primarily non-specific it would explain why there are no significant observed differences between the c-MYC-ZipΔDT and c-MYC-Zip peptides, with respect to inhibition of the leucine zipper conformation. For both heterodimers we observe a loss of the leucine zipper conformation upon the addition of 10058-F4; however, for c-MYC-Zip we also see a compaction of the $[C+4H]^{4+}$ charge state. This is coupled with a shift in charge state, a decrease in the $[C+5H]^{5+}$ species, paired with an increase in the $[C+4H]^{4+}$ state, observed for both c-MYC peptides, and suggests that the more loose form seen in the absence of the ligand and as the $[C+5H]^{5+}$ species may be *en route* to zipper formation.

The results presented here highlight that solution-based methods such as CD, for quantifying small changes in secondary structure for complex, dynamic, principally disordered systems provides ambiguous results and cannot provide the clear distinction between components that MS and DT IM-MS has achieved here. In addition to separating out ions of different *m/z* enabling monomers, homodimers and heterodimers to be clearly distinguished the conformational families of each can be individually determined.

These results show that the c-MYC:MAX heterodimer comprising the leucine zipper peptide regions can exist in two distinct conformations, one of which is compact and hence likely disordered and a second which corresponds to an extended and potentially 'coiled-coil' structure. Using DT IM-MS these different conformational families can be identified and both distinct forms studied simultaneously. The formation of this leucine zipper coiled-coil type structure is inhibited by the ligand 10058-F4, pushing the equilibrium towards the disordered and supposedly inactive form, supporting that this ligand stabilises the disordered form. This proof of concept study has provided clear evidence that the combination of MS and DT IM-MS could provide a useful ligand screening method in the development of compounds designed to disrupt formation of the c-MYC:MAX leucine zipper. Furthermore, DT IM-MS is a powerful tool to distinguish between multiple conformations of dynamic peptide systems, which can be difficult to distinguish in traditional solution-phase studies[61].

3.5 References

(1) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L., *Curr. Opin. Struct. Biol.*, **2008**, *18*, 756-764.
(2) Eliezer, D., *Curr. Opin. Struct. Biol.*, **2009**, *19*, 23-30.
(3) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z., *J. Mol. Graph. Model.*, **2001**, *19*, 26-59.
(4) Mendoza-Espinosa, P.; García-González, V.; Moreno, A.; Castillo, R.; Mas-Oliva, J., *Mol. Cell. Biochem.*, **2009**, *330*, 105-120.
(5) Lacy, E. R.; Filippov, I.; Lewis, W. S.; Otieno, S.; Xiao, L.; Weiss, S.; Hengst, L.; Kriwacki, R. W., *Nat. Struct. Mol. Biol.*, **2004**, *11*, 358-364.
(6) Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E., *Nature*, **2002**, *415*, 549-553.
(7) Wright, P. E.; Dyson, H. J., *Curr. Opin. Struct. Biol.*, **2009**, *19*, 31-38.
(8) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K., *Annu Rev Biophys*, **2008**, *37*, 215-246.
(9) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P., *J. Mol. Biol.*, **2004**, *338*, 1015-1026.
(10) Keppel, T. R.; Howard, B. A.; Weis, D. D., *Biochemistry*, **2011**, *50*, 8722-8732.
(11) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K., *J. Mol. Biol.*, **2002**, *323*, 573-584.
(12) Metallo, S. J., *Curr. Opin. Cell. Biol.*, **2010**, *14*, 481-488.
(13) Wang, J.; Cao, Z.; Zhao, L.; Li, S., *Int. J. Mol. Sci.*, **2011**, *12*, 3205-3219.
(14) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J., *Curr. Opin. Struct. Biol.*, **2011**, *21*, 432-440.
(15) Dunker, A. K.; Uversky, V. N., *Curr. Opin. Pharmacol.*, **2010**, *10*, 782-788.
(16) Cheng, Y.; LeGall, T.; Oldfield, C. J.; Mueller, J. P.; Van, Y.-Y. J.; Romero, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K., *Trends Biotechnol.*, **2006**, *24*, 435-442.
(17) Hammoudeh, D. I.; Follis, A. V.; Prochownik, E. V.; Metallo, S. J., *J. Am. Chem. Soc.*, **2009**, *131*, 7390-7401.
(18) Fernandez, P. C.; Frank, S. R.; Wang, L.; Schroeder, M.; Liu, S.; Greene, J.; Cocito, A.; Amati, B., *Genes Dev.*, **2003**, *17*, 1115-1129.
(19) Li, Z.; Van Calcar, S.; Qu, C.; Cavenee, W. K.; Zhang, M. Q.; Ren, B., *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, *100*, 8164-8169.
(20) Lavigne, P.; Crump, M. P.; Gagne, S. M.; Hodges, R. S.; Kay, C. M.; Sykes, B. D., *J. Mol. Biol.*, **1998**, *281*, 165-181.
(21) Vita, M.; Henriksson, M. In *Semin Cancer Biol*; Elsevier: 2006; Vol. 16, p 318-330.
(22) Follis, A. V.; Hammoudeh, D. I.; Wang, H.; Prochownik, E. V.; Metallo, S. J., *Chem. Biol.*, **2008**, *15*, 1149-1155.
(23) Wang, H.; Hammoudeh, D. I.; Follis, A. V.; Reese, B. E.; Lazo, J. S.; Metallo, S. J.; Prochownik, E. V., *Mol. Cancer Ther.*, **2007**, *6*, 2399-2408.
(24) Lavigne, P.; Kondejewski, L. H.; Houston Jr, M. E.; Sönnichsen, F. D.; Lix, B.; Sykes, B. D.; Hodges, R. S.; Kay, C. M., *J. Mol. Biol.*, **1995**, *254*, 505-520.
(25) Mittag, T.; Forman-Kay, J. D., *Curr. Opin. Struct. Biol.*, **2007**, *17*, 3-14.
(26) Clore, G. M.; Gronenborn, A. M., *Trends Biotechnol.*, **1998**, *16*, 22-34.
(27) Congreve, M.; Murray, C. W.; Blundell, T. L., *Drug Discov. Today*, **2005**, *10*, 895-907.
(28) Beveridge, R.; Chappuis, Q.; Macphee, C.; Barran, P., *Analyst*, **2013**, *138*, 32-42.
(29) Maurizio, E.; Cravello, L.; Brady, L.; Spolaore, B.; Arnoldo, L.; Giancotti, V.; Manfioletti, G.; Sgarra, R., *J. Proteome. Res.*, **2011**, *10*, 3283-3291.
(30) Zhang, Z.; Smith, D. L., *Protein Sci.*, **1993**, *2*, 522-531.
(31) Chowdhury, S. K.; Katta, V.; Chait, B. T., *J. Am. Chem. Soc.*, **1990**, *112*, 9012-9013.
(32) Konermann, L.; Douglas, D. J., *Biochemistry*, **1997**, *36*, 12296-12302.
(33) Frimpong, A. K.; Abzalimov, R. R.; Uversky, V. N.; Kaltashov, I. A., *Proteins: Struct. Funct. Bioinf.*, **2010**, *78*, 714-722.
(34) Iakoucheva, L. M.; Kimzey, A. L.; Masselon, C. D.; Bruce, J. E.; Garner, E. C.; Brown, C. J.; Dunker, A. K.; Smith, R. D.; Ackerman, E. J., *Protein Sci,*, **2001**, *10*, 560-571.

(35) Brocca, S.; Testa, L.; Sobott, F.; Samalikova, M.; Natalello, A.; Papaleo, E.; Lotti, M.; De Gioia, L.; Doglia, S. M.; Alberghina, L.; Grandori, R., *Biophys. J.*, **2011**, *100*, 2243-2252.
(36) Bernstein, S. L.; Liu, D.; Wyttenbach, T.; Bowers, M. T.; Lee, J. C.; Gray, H. B.; Winkler, J. R., *J. Am. Soc. Mass Spectrom.*, **2004**, *15*, 1435-1443.
(37) Smith, D. P.; Radford, S. E.; Ashcroft, A. E., *Proc. Natl. Acad. Sci. U. S. A.*, **2010**, *107*, 6794-6798.
(38) Shimotakahara, S.; Shiroyama, Y.; Fujimoto, T.; Akai, M.; Onoue, T.; Seki, H.; Kado, S.; Machinami, T.; Shibusawa, Y.; Uéda, K., *J. Biophys. Chem.*, **2012**, *3*, 149-155.
(39) Natalello, A.; Benetti, F.; Doglia, S. M.; Legname, G.; Grandori, R., *Proteins*, **2011**, *79*, 611-621.
(40) Ly, T.; Julian, R. R., *J. Am. Soc. Mass Spectrom.*, **2008**, *19*, 1663-1672.
(41) Grabenauer, M.; Bernstein, S. L.; Lee, J. C.; Wyttenbach, T.; Dupuis, N. F.; Gray, H. B.; Winkler, J. R.; Bowers, M. T., *J. Phys. Chem. B*, **2008**, *112*, 11147-11154.
(42) Illes-Toth, E.; Dalton, C.; Smith, D., *J. Am. Soc. Mass Spectrom.*, **2013**, *24*, 1346-1354.
(43) Harvey, S. R.; Porrini, M.; Stachl, C.; MacMillan, D.; Zinzalla, G.; Barran, P. E., *J. Am. Chem. Soc.*, **2012**, *134*, 19384-19392.
(44) Dang, C. V.; McGuire, M.; Buckmire, M.; Lee, W. M., *Nature*, **1989**, *337*, 664-666.
(45) Brownlie, P.; Ceska, T.; Lamers, M.; Romier, C.; Stier, G.; Teo, H.; Suck, D., *Structure*, **1997**, *5*, 509-520.
(46) Sauve, S.; Tremblay, L.; Lavigne, P., *J. Mol. Biol.*, **2004**, *342*, 813-832.
(47) Whitmore, L.; Wallace, B. A., *Nucleic Acids Res.*, **2004**, *32*, W668-W673.
(48) Whitmore, L.; Wallace, B. A., *Biopolymers*, **2008**, *89*, 392-400.
(49) Johnson, W. C., *Proteins-Structure Function and Genetics*, **1999**, *35*, 307-312.
(50) Muhle-Goll, C.; Gibson, T.; Schuck, P.; Schubert, D.; Nalis, D.; Nilges, M.; Pastore, A., *Biochemistry*, **1994**, *33*, 11296-11306.
(51) Yin, S.; Xie, Y.; Loo, J. A., *J. Am. Soc. Mass Spectrom.*, **2008**, *19*, 1199-1208.
(52) Hopper, J. T. S.; Sokratous, K.; Oldham, N. J., *Anal. Biochem.*, **2012**, *421*, 788-790.
(53) Liu, L.; Michelsen, K.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S., *J. Am. Chem. Soc.*, **2010**, *132*, 17658-17660.
(54) Liu, L.; Kitova, E. N.; Klassen, J. S., *J. Am. Soc. Mass. Spectrom.*, **2011**, *22*, 310-318.
(55) Michel, J.; Cuchillo, R., *PLoS ONE*, **2012**, *7*, e41070.
(56) Kabsch, W.; Sander, C., *Biopolymers*, **1983**, *22*, 2577-2637.
(57) Barran, P. E.; Polfer, N. C.; Campopiano, D. J.; Clarke, D. J.; Langridge-Smith, P. R. R.; Langley, R. J.; Govan, J. R. W.; Maxwell, A.; Dorin, J. R.; Millar, R. P.; Bowers, M. T., *Int. J. Mass. Spectrom.*, **2005**, *240*, 273-284.
(58) Breuker, K.; McLafferty, F. W., *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105*, 18145-18152.
(59) Barran, P. E., *Angew. Chem.*, **2011**, *50*, 3120-3122.
(60) Hoaglund-Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E., *Chem. Rev.*, **1999**, *99*, 3037-3080.
(61) Murzin, A. G., *Science*, **2008**, *320*, 1725-1726.

4

Investigating conformational equilibrium: Lymphotactin

The theory that a protein will exist in a single, unique, evolutionary constrained conformation has been called into question by the characterisation of metamorphic proteins, which exist in multiple functional folds. Successful transmission of proteins into the gas-phase does not depend on protein fold and therefore mass spectrometry is inherently well suited to study conformationally dynamic systems. This chapter presents the combination of MS, DT IM-MS and ECD as a tool to study proteins which are dynamic with respect to conformation, considering the metamorphic protein lymphotactin and specific mutants designed to stabilise both distinct conformations.

4.1 Introduction

Lymphotactin (Ltn) is a unique chemokine for many reasons, for example, it contains an extended C-terminal sequence which forms an intrinsically disordered (ID) tail; it also contains a single disulfide bond as opposed to the two normally found in chemokines[1,2]. Most significantly Ltn is a metamorphic protein capable of adopting two distinct folds, for the same sequence of amino acids[3], existing in a monomeric and dimeric fold (Chapter 1 section 1.3.3). Lymphotactin exists in equilibrium between a monomeric conserved chemokine fold (Ltn10) and a dimeric fold (Ltn40) unique amongst the chemokine subclass (Figure 4.1).

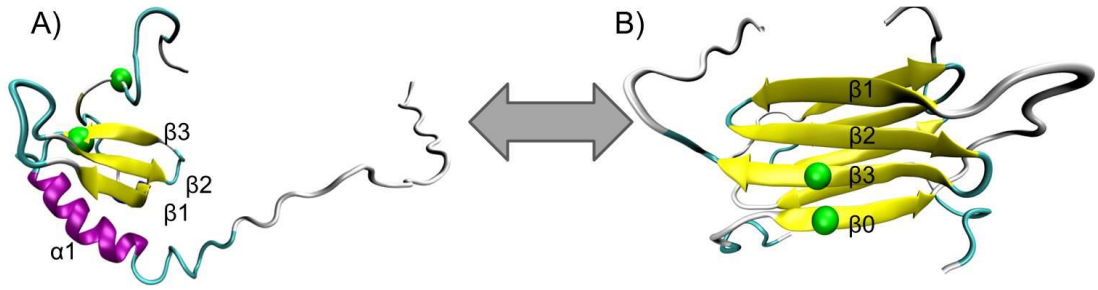

Figure 4.1: A) Monomeric Ltn10 (PDB 1J9O), B) Dimeric Ltn40 (PDB 2JP1). Green spheres highlight position of the disulfide bond.

A number of proteins which can adopt multiple, distinct conformations have been previously identified, however, these proteins were initially considered 'special cases' with conformational switching often being observed to be triggered by environmental factors such as pH-induced transitions, redox-controlled transitions and disulfide isomerisation[4-7]. The discovery of the metamorphic proteins Mad2 and Ltn, which can adopt distinct conformations in equilibrium in a ligand-free state, however, called in to question not only the theory that a protein will exist in a single unique conformation but also if traditional structural elucidation techniques were inadvertently selecting against detection of multiple conformations[3,8].

Ltn is a well-studied metamorphic protein; both conformations adopted have been solved, conditions under which the conformational equilibrium can be shifted are known and biological functions of both folds are under study[1,3]. Ltn is therefore an ideal model metamorphic protein to study in order to determine if both distinct folds of such a protein can be preserved, transmitted and distinguished in the gas-phase[9]. Such a study would help determine if a mass spectrometric framework would be a suitable alternative to traditional structure elucidation techniques. Furthermore, Ltn10 is an excellent model system to probe protein stability in the gas-phase, because it adopts all major types of secondary structure;

containing an α-helix and three β-strands stabilised by a disulfide bond, as well as possessing an ID region.

The ability of MS-based techniques to study both structured and unstructured proteins, combined with the ability of IM-MS to distinguish between multiple conformations, even when they present at the same charge state, suggests that a MS-based approach could overcome the (unintentional) bias associated with many traditional techniques that consider and solve for only a single conformation. One particular advantage of MS-based techniques is they provide an opportunity to distinguish between intrinsic and extrinsic factors contributing to protein fold stability. Prior to performing gas-phase structural studies, however, one has to consider the effect of ionising and desolvating protein structures. It appears probable that protein ions generally undergo some structural collapse of side-chain groups within picoseconds of dehydration[10], however, onset of gross structural rearrangements can require tens of milliseconds to seconds, therefore one could propose that as long as experimental trapping times are kept short the solution fold may be retained[10-12].

MS-based techniques can be used to sample dynamic conformations as well as to take 'snapshots' of protein folding and unfolding, using labelling or label-free methods. Labelling methods can involve hydrogen/deuterium exchange or more recently a technique known as fast photochemical oxidation of proteins, both of which have produced some elegant studies of protein folding and dynamics in addition to distinguishing between protein conformations[13-19]. These techniques, however, generally involve digesting the sample prior to MS analysis and therefore conformational information obtained is an average of all species present in solution. Label-free MS-based methods can be used to consider all conformations present in solution individually, and generally have been focused on studying unfolding either *via* increasing the time ions spend in the gas-phase or through consideration of multiple charge states, and most frequently in combination with IM-MS studies[20-22].

IM-MS has been used to distinguish between distinct conformations of proteins and aggregating peptide systems which can undergo conformational switching, highlighting its applicability to investigate conformationally dynamic systems[23,24]. One such study presented by Jenner *et al*[25], studied NADPH-cytochrome p450 reductase (CPR), using MS and IM-MS. CPR exists in an 'open' and 'closed' conformation, where in its 'closed' conformation it is inaccessible to the large cytochrome p450 molecule and hence movement of domains is functionally essential. The IM-MS studies presented in this study not only highlighted that the two distinct conformations could be preserved into the gas-phase but also the population of each could be influenced by ionic strength of the solution prior to desolvation.

Furthermore, careful consideration of conformations, along with elegant mutagenesis studies, identified two key salt bridges essential for the compact conformations to form. Their findings highlighted that these stabilising interactions could be preserved in the gas-phase over the time scale of these experiments and that IM-MS can be a useful tool to study conformationally dynamic systems, allowing multiple conformations to be studied in each experiment.

Recently a study was presented by Skinner *et al*[26] which compared structural analysis from electron capture dissociation (ECD) experiments, to previously obtained results from DT IM-MS studies from the Bowers[27] and Clemmer[12] groups. This study focused on the unfolding of ubiquitin $[M+7H]^{7+}$ and highlighted that comparison of ECD and DT IM-MS has great promise to gain insight into unfolding of protein ions as a function of time spent in the gas-phase.

Although many MS fragmentation techniques exist, ECD is particularly advantageous for protein structural studies. This is due to the fast, non-ergodic, fragmentation process which occurs during ECD, in which bond dissociation following activation occurs much faster than typical bond vibration[28,29]. ECD is therefore thought not to perturb the higher order structure of proteins, enabling examination of the conformations adopted by a single protein over multiple charge states. Over the last couple of decades ECD has been used in a number of elegant studies of protein fold providing further information on protein conformations and their stabilities in the gas-phase[30-34].

Despite the obvious promise in the combination of these techniques[35,36], to-date a study combining these techniques to explore protein unfolding as a function of charge has not yet been presented. Furthermore, the applicability of such a combination of techniques to distinguish between multiple conformations of the same protein has not been addressed. The combination of these techniques would allow the study of fold and unfolding by IM-MS, with amino acid detail provided by ECD analysis, and could be used to delineate the stability of various regions of protein fold. This chapter focuses on the application of these techniques to differentiate between two distinct folds of wild type (WT) Ltn. Studies are performed on mutants of Ltn designed to stabilise each distinct conformation, in order to further probe these two folds and assess the extent to which these mutants do indeed favour their designed fold.

4.2 Experimental

4.2.1 Protein samples

All recombinant human lympohactin (Ltn) samples were expressed and purified by the Volkman Research Group (Medical College of Wisconsin, Milwaukee, USA) as previously described[1,37,38]. Due to differences in expression and purification systems used for WT Ltn and the mutants CC3 and W55D, residues M63 and M73 were mutated to valine and alanine respectively in both CC3 and W55D. These mutations cause no distinguishable difference in structure or function[3,38]. Protein samples were received as lyophilised samples and stored at -20 °C for subsequent study. Stock solutions were typically prepared at 100, 150 or 200 μM in high purity water or 20 mM ammonium acetate (AmAc) and stored in aliquots at -20 °C until use. Concentrations of protein solutions were verified using the Pierce BCA concentration assay (Thermo Fischer Scientific, Wilmington, DE, USA) following manufacturer's instructions, using volumes 10 times lower than stated in the manual, afforded through use of low volume UV cuvettes.

4.2.2 Sample preparation for MS studies

All MS studies into the effect of solution conditions on the monomer-dimer equilibrium of wild type (WT) and W55D Ltn constructs were performed on 50 μM protein solutions, prepared in high purity water or AmAc at 20, 50, 100 or 150 mM concentration.

To study the effect of pH on the equilibrium of WT Ltn 20 mM AmAc was used as buffer, the pH of which was altered over the range 5.8 to 2.8 through drop-wise addition of formic acid (VWR International Ltd, UK). The effect on the WT Ltn monomer:dimer equilibrium of adding NaCl (Sigma Aldrich Ltd, UK) was studied using 20 mM AmAc plus either 1 or 5 mM NaCl. W55D pH studies were performed in 20 mM AmAc buffer, over the pH range 9.8-2.8, buffer pH was adjusted through drop-wise addition of either ammonia (Fisher Scientific, UK) or formic acid. Monomer-dimer equilibrium for CC3 and WT 1-72 were studied using 50 μM protein solutions in 20 mM AmAc at pH 6.8.

The ability of the structural core to bind the ID tail was probed using a sample solution containing a 1:1 mixture of WT 1-72 plus WT 72-93 each at a concentration of 50 μM in 20 mM AmAc.

All MS experiments were performed on a Q-TOF Ultima (Waters, Manchester, UK), for typical operating conditions see Chapter 2 section 2.2. All studies, except those involving WT 72-93, were performed on the Q-TOF Ultima prior to the high mass upgrade.

4.2.3 Sample preparation for DT IM-MS studies

For CCS determination all Ltn samples were prepared at 100 μM in 20 mM AmAc. Additionally, CC3 supercharged samples were prepared at 100 μM in 20 mM AmAc plus 1 % *meta*-nitrobenzyl alcohol (*m*-NBA), purchased from Sigma Aldrich Company Ltd (Dorset, UK). DT IM-MS measurements were performed on an in-house modified Q-TOF (Micromass UK Ltd.), as described in Chapter 2 Section 2.3.

4.2.4 Sample preparation for ECD studies

Electron capture dissociation (ECD) fragmentation studies for monomeric WT Ltn and WT 1-72 were performed on solutions of 5 μM protein in 100 mM AmAc. ECD studies for CC3, W55D and dimeric WT Ltn, were performed on 30 μM protein in 100 mM AmAc. CC3 supercharged samples were prepared at 30 μM in 20 mM AmAc plus 1 % *m*-NBA.

All ECD studies were performed on a 12T Apex Ultra Qh FT-ICR mass spectrometer (Bruker Daltonik Gmbh, Bremen, Germany), instrumental details can be found in Chapter 2 section 2.5.

4.2.5 Molecular modelling

Ltn molecular modelling was performed by Dr Massimiliano Porrini. In brief, theoretical CCS were determined for Ltn10, Ltn40 and CC3 Ltn from the NMR structures (PDB identifiers 1J9O, 2JP1 and 2HDM respectively) plus Ltn10 1-72. The structures were minimised *in vacuo* using the *sander* engine of Amber11[39] and implementing Amber99SB-ILDN[40] prior to determining their CCS using the trajectory method of MOBCAL[41]. For Ltn40, the PDB file (2JP1) contains 20 candidate structures and theoretical CCS were determined for each using the above procedure and the average value reported as the theoretical CCS. Details on additional molecular dynamics simulations performed on WT Ltn, to study salt bridges present within the gas-phase structures, can be found in Appendix 3 section A3.1.

4.3 Results and discussion

4.3.1 Studying the conformational equilibria of WT lymphotactin using MS

The initial stage of investigation into this metamorphic protein focused on determining if both distinct forms, *ie* monomer (Ltn10) and dimer (Ltn40), could be preserved and transferred into the gas-phase.

The mass spectrum for wild type (WT) Ltn was acquired in 20 mM ammonium acetate (AmAc), Figure 4.2. Both monomeric (M_{WT}) and dimeric (D_{WT}) species are observed in the spectrum, with monomeric species being more intense. Monomeric WT Ltn presents over a range of charge states from $[M_{WT}+4H]^{4+}$ to $[M_{WT}+10H]^{10+}$, this suggests the protein is dynamic and flexible. Interestingly, there are two species ($[M_{WT}+6H]^{6+}$ and $[M_{WT}+7H]^{7+}$) which are particularly intense suggesting the conformations of Ltn which can accept either six or seven net protons are particularly favourable. Dimeric WT Ltn is also observed, over the charge state range $[D_{WT}+8H]^{8+}$ to $[D_{WT}+11H]^{11+}$, at lower intensity than the monomer. This indicates both distinct forms of WT Ltn can be preserved and transmitted into the gas-phase, however, here the equilibrium between the two is pushed strongly towards monomer.

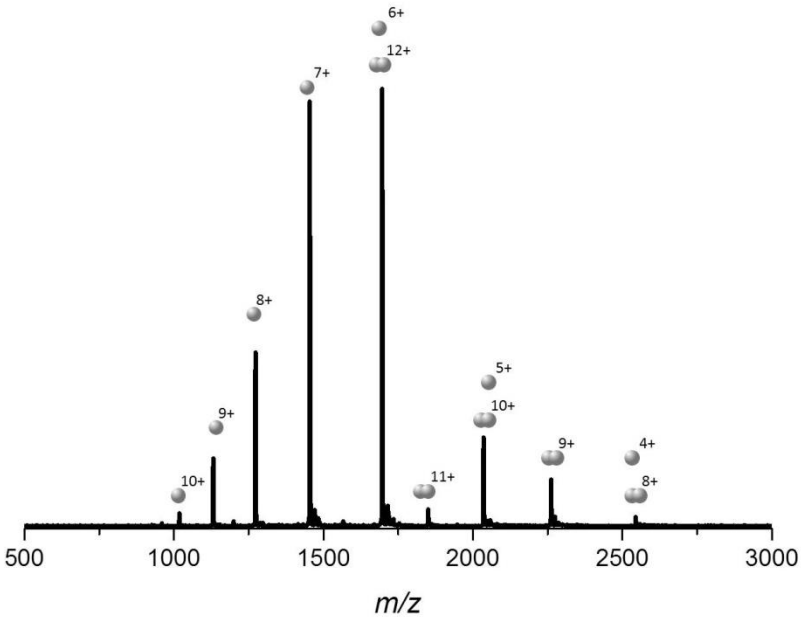

Figure 4.2: The mass spectrum obtained for 50 μM WT Ltn in 20 mM AmAc pH 6.8.

NMR studies have shown that the solution equilibrium can be altered through the choice of specific solvent conditions; Ltn10 was found to be favoured in 20 mM phosphate buffer with 200 mM NaCl (pH 6) and Ltn40 was solved in 20 mM sodium phosphate (pH 6). It has been speculated that charge repulsion between the R23 and R43 residues destabilises the

chemokine (Ltn10) fold and promotes formation of the dimeric (Ltn40) fold[2]. This repulsion can be overcome by binding a chloride which will stabilise Ltn10 by neutralising the repulsive effect, hence high salt content pushes the equilibrium towards Ltn10. High concentrations of salts such as NaCl, however, are not compatible with MS-based studies, as they ionise readily and are transmitted easily, generally suppressing protein signal. In addition, salts can bind to protein and depending on their volatility can be retained after ionisation and desolvation. This causes an apparent diminished intensity for the protein signal as the observed signal is split over several different species, containing different numbers of salts adducted to protein, as shown in Appendix 3 Figure A3.1 for WT Ltn in 20 mM AmAc plus 5 mM NaCl. Therefore it is not feasible to directly copy the solution conditions employed in NMR studies of Ltn to MS studies.

There are solvent conditions that can be altered whilst remaining compatible with n-ESI-MS studies; for example the strength of volatile buffer as well as solution pH. The effect of altering both of these and their influence on Ltn monomer-dimer equilibrium were considered. Figure 4.3A illustrates the effect of altering buffer concentration from 0 mM AmAc to 100 mM AmAc. For WT Ltn in water (Figure 4.3A top) only monomeric species are observed and a major shift in the charge-state distribution (CSD) to higher charge states is observed, as compared with WT Ltn in 10-100 mM AmAc. These low *m/z* values are expected, as water, often in combination with organic solvents, promotes high charge states, whilst addition of volatile buffer salts is known to lower the charge states observed[42]. The difference between water and 10 mM AmAc is dramatic and highlights that for Ltn the obtained spectra can be strongly influenced by solvent conditions. Furthermore water is not a suitable solvent to probe the Ltn10 and Ltn40 folds and further studies require buffered 'near-native' solutions.

The mass spectra obtained for Ltn in 10-100 mM AmAc highlight that under all buffered conditions studied here monomeric and dimeric forms of WT Ltn are preserved. It is clear that as buffer strength is increased the CSD narrows and shifts to lower charge states, as expected. The shift in charge state over this range of buffer conditions is, however, very small, suggesting that Ltn adopts similar states under all AmAc buffer strengths studied. At high buffer strengths it is possible to slightly increase the relative proportion of dimer, however, a major shift in the equilibrium is not observed and monomer is always more intense. The mass spectrum obtained in 20 mM AmAc presents a wide range of charge states for both monomer and dimer, enabling a range of species to be considered and hence all

other studies of WT Ltn (except ECD studies, see section 4.2.4) were carried out at this buffer strength.

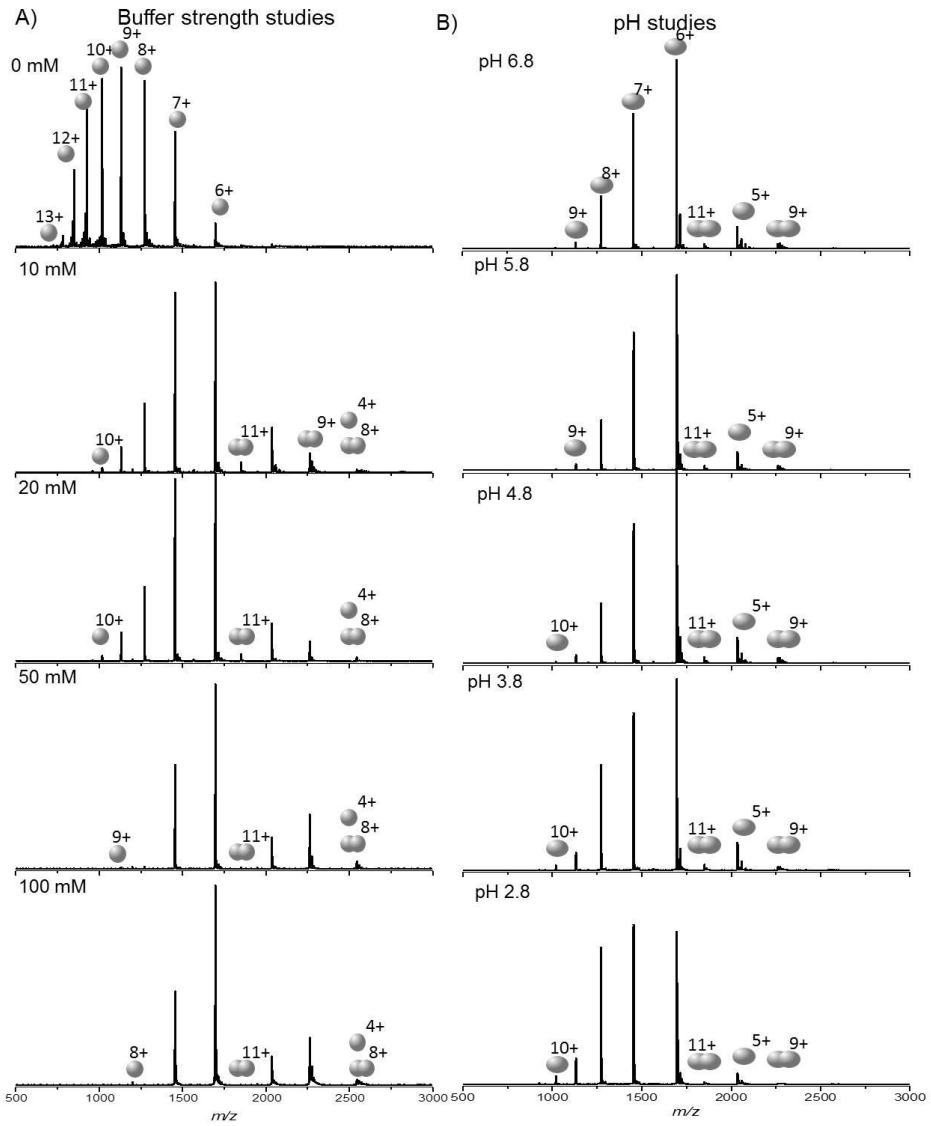

Figure 4.3: WT Ltn mass spectra for 50 µM Ltn A) Buffer strength, 0-100 mM AmAc, B) Buffer pH, 6.8-2.8, altered through addition of formic acid to 20 mM AmAc. For top panels all major species are labelled, in all other panels the highest and lowest charge states of monomer and dimer are labelled.

All NMR studies on this metamorphic protein were performed at pH 6, however, 20 mM AmAc with no additional additives is pH 6.8. Therefore in order to determine if studies would be better performed at a lower pH, in addition to probing the effect of pH on monomer-dimer equilibrium, the pH of 20 mM AmAc buffer was lowered through addition of formic acid (Figure 4.3B). Ltn is extremely pH stable, with no observed change in the CSD of both monomer and dimer between pH 6.8 and 4.8. At pH 3.8 a slight shift in the CSD is observed for monomeric Ltn. $[M_{WT}+6H]^{6+}$ is still observed to be the most intense

species, however, there is a marginal increase in the intensity of higher charge states. At pH 2.8 an additional shift in the CSD is detected with $[M_{WT}+7H]^{7+}$ now being the most intense species. Furthermore, at pH 2.8 the intensity of dimeric species decreases. The changes observed in the CSDs, however, are much less dramatic than would be expected for a typical protein[43], highlighting the stability of WT Ltn, enabling MS experiments to proceed at pH 6.8, which simplifies the solution conditions.

Interestingly under all MS compatible solution conditions studied here monomeric Ltn is consistently observed to be the dominant form, suggesting under these conditions it is more stable than the dimer, or under these conditions a proportion of the dimer is dissociating to monomer increasing the observed monomeric signal, a hypothesis which can be explored using DT IM-MS.

4.3.2 Examining the metamorphic conformations of WT Ltn by DT IM-MS and ECD

4.3.2.1 DT IM-MS of monomeric WT Ltn

CCS for monomeric WT Ltn were determined over the charge state range $[M_{WT}+4H]^{4+}$ to $[M_{WT}+9H]^{9+}$ (Figure 4.4 and Appendix 3 Table A3.1). As the charge increases, the CCS increase can be attributed to increased access to protonatable sites in solution for the more unfolded forms and to Coulombic effects. The lowest charged species observed ($[M_{WT}+4H]^{4+}$) is present at low intensity with a CCS significantly smaller than that of $[M_{WT}+5H]^{5+}$; which is assigned to a significantly collapsed species, as has been reported previously for gas-phase monomeric proteins[11,20], suggesting that this form is not structurally stable.

MS indicated the most significantly populated species observed under these conditions presented with between five and seven net protons (Figure 4.2). Interestingly from DT IM-MS it is observed that these species all present with exceptionally similar CCS (~1000 $Å^2$), establishing that there is a compact, stable conformational family in solution, which can accept between five and seven net protons without a major unfolding transition. The slight increase in CCS observed over this charge state range can be ascribed to Coulombic effects. A significant increase (up to 42 % larger than $[M_{WT}+7H]^{7+}$) in CCS is observed for $[M_{WT}+8H]^{8+}$ and $[M_{WT}+9H]^{9+}$, suggesting a key unfolding event has occurred.

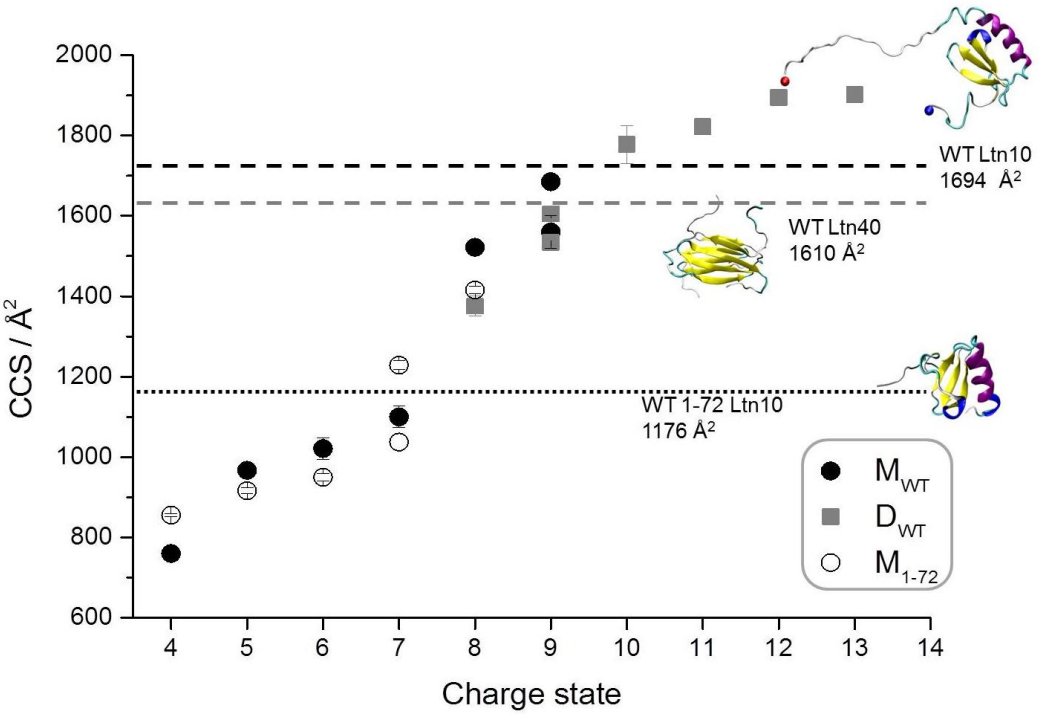

Figure 4.4: Average CCS from three repeats determined for WT Ltn, error bars show standard deviation between repeats. WT monomer species are represented by black dots and dimer species by grey squares. WT 1-72 monomeric CCS are represented by open circles. Theoretical CCS for WT Ltn10 (PDB 1J9O) and Ltn40 (PDB 2JP1) are shown by dashed black and grey lines respectively. Dotted line represents the theoretical CCS obtained for the clipped structure of PDB file 1J9O, removing residues 73-93.

From DT IM-MS it is clear that here monomeric WT Ltn exists in three distinct conformational families; collapsed ($[M_{WT}+4H]^{4+}$), compact ($[M_{WT}+5H]^{5+}$ to $[M_{WT}+7H]^{7+}$) and extended ($[M_{WT}+8H]^{8+}$ and $[M_{WT}+9H]^{9+}$). In order to determine if any of these conformational families could be due to WT Ltn in its solution-phase conformation (Ltn10) a comparison between experimental CCS and predicted CCS from the solved structure (PDB 1J9O) was made (Figure 4.4). The theoretical CCS determined from the NMR structure was 1694 Å^2 which is significantly higher than the $[M_{WT}+4H]^{4+}$ to $[M_{WT}+7H]^{7+}$ species but compares favourably with the CCS determined for extended conformational family. PDB 1J9O contains the ID tail extended from the structural core, resulting in a large theoretical CCS and suggesting that the extended conformational family retains the Ltn10 fold, with the ID tail unfolded from the structural core. It is important to note that $[M_{WT}+9H]^{9+}$ is observed experimentally to exist in two conformations the second of which is even more extended (up to 50 % larger than $[M_{WT}+7H]^{7+}$) and corresponds to a conformer further along an unfolding pathway, a point discussed in more detail below.

The theoretical CCS for Ltn10 is significantly larger than the compact conformational family. A second theoretical CCS was then determined for the N-terminus and structural

core of Ltn without the ID tail, achieved by clipping the NMR structure (PDB 1J9O) removing residues 73-93. When the tail is removed a theoretical CCS of 1176 $Å^2$ is obtained, which is in agreement with the experimental CCS obtained for full length WT Ltn in the charge state range $[M_{WT}+5H]^{5+}$ to $[M_{WT}+7H]^{7+}$, suggesting that these species retain the Ltn10 fold with the tail associated or wrapped around the structural core.

To study this hypothesis an Ltn construct was considered which consists only of the N-terminus and structural core (residues 1-72) known as WT 1-72. Experimental CCS were determined for monomeric WT 1-72 (M_{1-72}) (Figure 4.4 open circles and Appendix 3 Table 3.2), and compare exceptionally favourably with full length M_{WT} Ltn in its compact conformation. It was therefore surmised that in its compact conformational ensemble full length M_{WT} Ltn exists in the Ltn10 fold with the ID tail wrapped around or associated with the core. Interestingly, WT 1-72 also contains an extended conformational family, with a second larger conformation observed for $[M_{1-72}+7H]^{7+}$ in addition to an extended $[M_{1-72}+8H]^{8+}$ species, a point which is discussed in detail in section 4.3.2.4.

Finally, the ability of the structural core (WT 1-72) to bind the ID tail was studied experimentally by MS through the addition of a peptide consisting of the tail residues (WT 72-93) to a solution of WT 1-72. The spectra obtained from a 1:1 mixture confirm that the tail can bind the structural core (Appendix 3 Figure A3.2). Hence, from the studies of WT 1-72 it can be inferred that in its compact conformational family ($[M_{WT}+5H]^{5+}$ to $[M_{WT}+7H]^{7+}$) monomeric full length WT Ltn is most likely in the Ltn10 fold with the tail wrapped around the structural core. The tail unfolds to form the extended conformational family causing a dramatic increase in CCS at charge states 8 and 9+.

4.3.2.2 DT IM-MS of dimeric WT Ltn

The CCS of dimeric Ltn were determined over charge states $[D_{WT}+8H]^{8+}$ to $[D_{WT}+13H]^{13+}$ (Figure 4.4 and Appendix 3 Table A3.3). Dimeric Ltn species over the charge state range $[D_{WT}+10H]^{10+}$ to $[D_{WT}+13H]^{13+}$ have very similar CCS ~1778-1902 $Å^2$. These experimental values are compared to a theoretical CCS obtained from the NMR structure of Ltn40 and are found to be in good agreement but marginally larger than the predicted CCS (1610 $Å^2$). The PDB file (2JP1) used to determine the theoretical CCS, however, does not contain the full ID tail sequence, which would explain the slightly smaller predicted size. In fact this PDB file omits 33 amino acids from each chain, yet the difference between this theoretical value and experimental values (obtained from full length protein) is very small. This suggests that the

dimeric species are in the Ltn40 conformation but once again the ID tail is wrapped around or associated with the structural core, as observed for Ltn10.

The CCS for the lowest charge states, $[D_{WT}+8H]^{8+}$ and $[D_{WT}+9H]^{9+}$ are significantly smaller than the other dimer charge states (up to 29 % smaller than $[D_{WT}+10H]^{10+}$) and up to 15 % smaller than the predicted CCS, suggesting these species are either the product of 'gas-phase collapse' or a less structured solution-phase encounter complex between the unfolded monomeric chains. Additionally, two conformations are observed for $[D_{WT}+9H]^{9+}$, suggesting that this may be a transition species between a collapsed and structurally more stable form.

4.3.2.3 ECD of monomeric WT Ltn, Ltn10

From DT IM-MS experiments of WT Ltn, information on the fold and unfolding can be obtained. In order to further probe the fold, ECD fragmentation studies were then performed. In these studies higher buffer strength (100 mM AmAc) was implemented due to the harsher desolvation conditions of the FT ICR-MS instrument used here. The use of higher buffer strength enables a spectrum to be obtained (Appendix 3 Figure A3.3) which is similar to that of WT Ltn prepared in 20 mM AmAc and studied on the MS and DT IM-MS instruments; although a slightly higher range of charge states are observed here.

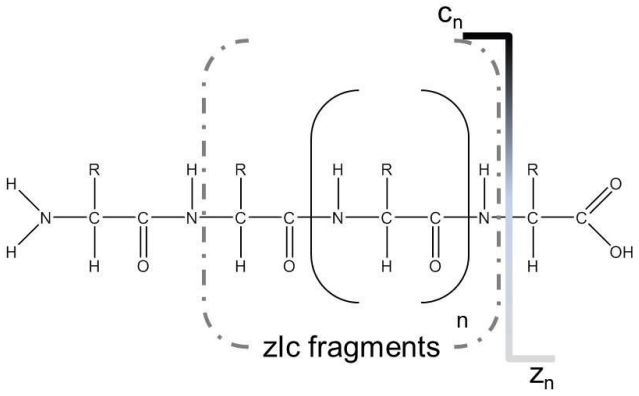

Figure 4.5: Schematic representation of zIc, c- and z-type fragments.

For ECD analysis of Ltn10, charge states $[M_{WT}+7H]^{7+}$ to $[M_{WT}+10H]^{10+}$ were observed at high enough intensity and each charge state was studied in turn. Fragments formed during this process are identified by considering fragmentation producing c- and z-type fragments[44] (Chapter 1 section 1.5.3.4), as well as internal fragments which are formed by multiple backbone cleavages resulting in internal stretches of amino acids being observed (Figure 4.5), which since they contain a 'z' fragment at their N-terminus and a 'c' fragment at their

C-terminus are denoted as ‘zIc’ fragments. Once all fragments had been identified their percentage yields were determined as a function of cleavage site (see Chapter 2 section 2.5.1) and are shown mapped onto the Ltn10 fold in Figure 4.6. In Figure 4.6 the fragmentation occurring for each individual charge state studied is plotted as percentage yield versus the fragmentation site, amino acid at which fragmentation occurs, and identifying the species as a zIc c- or z-type fragment, obtaining a fragmentation map for a single charge state. The stacking of the fragmentation maps for each charge state enable trends across charge states to be identified, as shown in Figure 4.6.

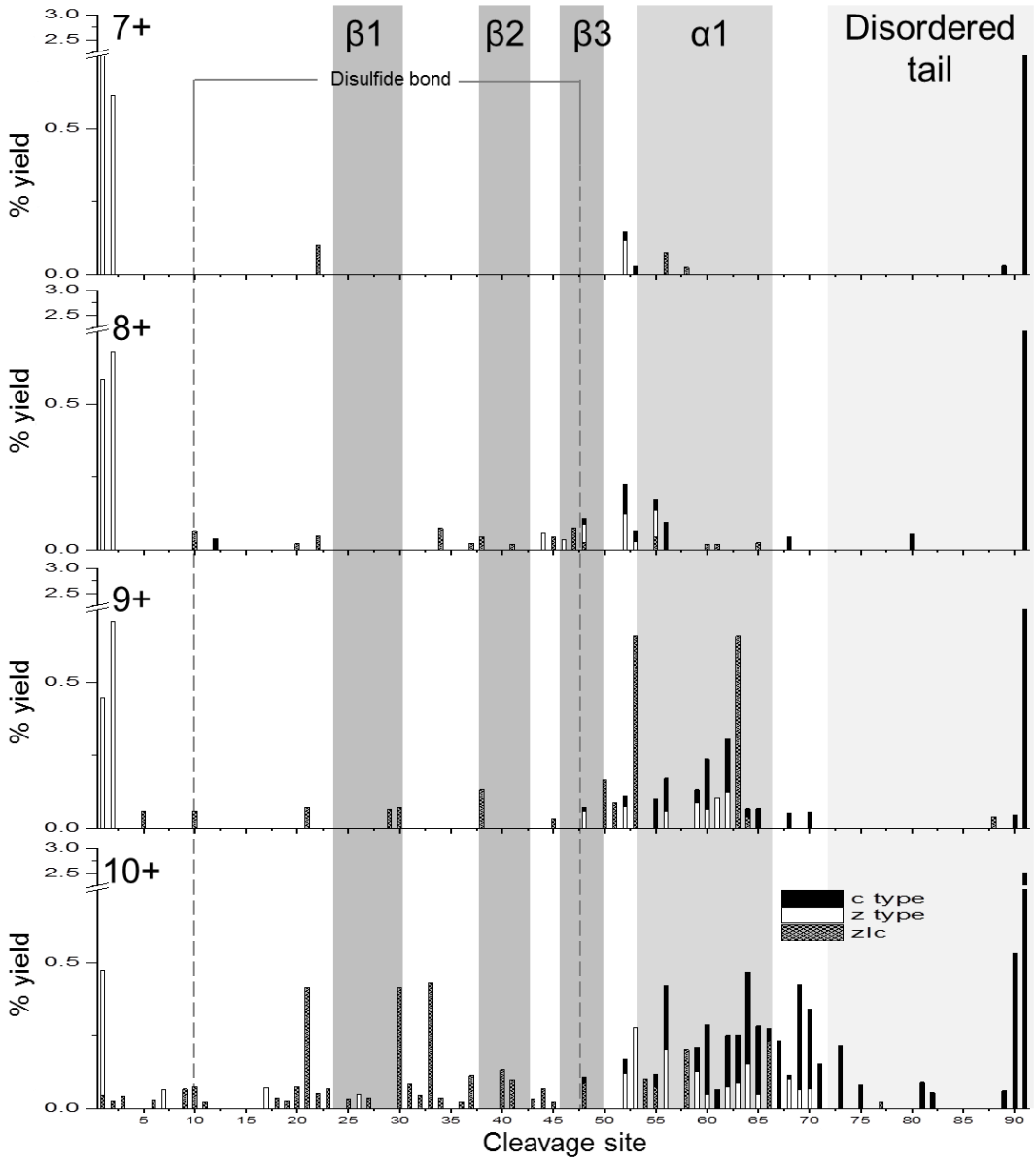

Figure 4.6: Percentage yields calculated for zIc, c- and z- type fragments as a function of cleavage site for monomeric Ltn $[M_{WT}+7H]^{7+}$ to $[M_{WT}+10H]^{10+}$. A y-axis scale break is used here to show low level fragmentation more clearly.

From the fragmentation map it is clear that as charge state increases there is a simultaneous increase in both the extent of backbone fragmentation and percentage yield of the fragment ions. This is expected due to a combination of effects, firstly ECD efficiency is known to be higher at higher charge states[45] and secondly at higher charge states species are more likely to be more unfolded, hence there are more available sites of cleavage. For all charge states the fragmentation yields are low, consistent with previous ECD studies[26,30], due to the high intensity of charge reduced species, which have undergone electron capture with no dissociation. Furthermore, Ltn contains a disulfide bond and disulfide bonds have been shown to be preferentially cleaved in ECD due to the high S-S affinity for $H^{\bullet}$ atoms[46], which would also result in a lower backbone fragmentation efficiency and, therefore, is an important consideration in such studies. Despite the low fragmentation yield insight into the protein fold and unfolding can be gleaned as fragmentation does not occur uniformly and observed fragments can be attributed to the presence (or absence) of stabilising non-covalent interactions present within the solution fold.

The lowest charge states studied by ECD ($[M_{WT}+7H]^{7+}$ and $[M_{WT}+8H]^{8+}$) exhibit very little fragmentation, apart from at the N- and C-termini, consistent with a compact structured conformational family stabilised by non-covalent interactions. As the charge state increases both the number of fragmentation sites and the fragmentation yield increases. This is most evident in the region which forms the α-helix in solution, in addition to the N- and C-terminus, consistent with these regions unfolding and hence being more amenable to cleavage, and in agreement with DT IM-MS results. Increased fragmentation is also observed in the regions between the β-sheet core and α-helix in addition to the region between the α-helix and the ID tail. Interestingly, at all charge states studied here very little fragmentation is observed within the intrinsically disordered tail region; attributed to the lack of backbone interactions in these regions, which are thought to be important for the ECD fragmentation mechanism[47]. For the $[M_{WT}+10H]^{10+}$ additional sites of fragmentation are observed along the ID tail, suggesting it may have undergone a reordering or restructuring event to enable fragmentation to occur.

For all charge states, the extent of backbone fragmentation and intensity of the fragmentation is low in the β-sheet core, often being observed only as zIc fragments as opposed to c- and z-type fragments. The low efficiency of fragmentation in this region is expected as the β-sheet core is stabilised by the disulfide bond, which would limit fragmentation in this region. Hence, when considering the fold and unfolding of a protein using gas-phase techniques it can be informative to consider interactions which would serve to stabilise the fold and their

potential to be retained in the gas-phase. The disulfide bond between residues 10-47 covalently bonds and stabilises these regions and several fragments are observed with this bridge intact (Figure 4.6 and Figure 4.7). Disulfide reduced fragments are also observed and increase in intensity with charge state, suggesting that this bond becomes more accessible to electron capture as Ltn unfolds. The disulfide bond would help to stabilise the β-sheet core and at all charge states minimal fragmentation occurs within this region. It is not until the highest charge state that fragmentation is observed in the β1 and β2 strands, suggesting Ltn10 has to be far along its unfolding pathways before these regions become destabilised. Indeed even at the highest charge state, in which the most extensive fragmentation is observed across the protein, only 25 % of the total fragments observed occurred due to fragmentation in the β-strands. When fragmentation does occur here it is of low intensity, only 17 % of the overall summed intensity of all observed fragments. Furthermore, cleavage in this region is most frequently a result of zIc fragmentation, with the zIc fragments generally consisting of more than one β-strand and suggesting these fragments partially retain some of the stabilising interactions present in the solution fold, attributed in part to the disulfide bond as well as non-covalent interactions.

Non-covalent interactions such as salt-bridges can be retained in the gas-phase and may strengthen in the solvent free environment[10]. In order to consider such interactions for Ltn molecular dynamics simulations (MD) were performed (Appendix 3 section A3.1). Firstly two separate 150 ns MD trajectories were run in explicit solvent to identify potential salt bridges in solution (Appendix 3 Table A3.4). An occupancy cut-off of 20 % was applied to identify the most likely salt bridges and pooling the two simulations resulted in identification of the following interactions; Glu3-Arg8, Asp6-Arg8, Glu30-Arg56, Asp57-Arg17 and Asp57-Arg60. In order to determine if these salt bridges can be retained in the gas-phase, *in vacuo* simulations were then run. In MS and DT IM-MS charge states 5-7+ are observed to be most intense hence these charge states are the most desirable targets for MD simulations, however, only charge states 7-10+ could be observed at high enough intensity to be subjected to ECD and therefore the 7+ charge state was chosen for MD studies. An *in vacuo* simulation of $[M_{WT}+7H]^{7+}$ determined the salt bridges between Glu3-Arg8, Asp57-Arg17 and Asp57-Arg64 all increase in occupancy in the gas-phase (Appendix 3 Table A3.5), furthermore, an additional salt bridge is formed between Asp63-Arg60. Both the Asp57-Arg60 and Asp63-Arg60 can act to stabilise the α-helix whilst Asp57-Arg17 would help position the α-helix within the structural core (Figure 4.7).

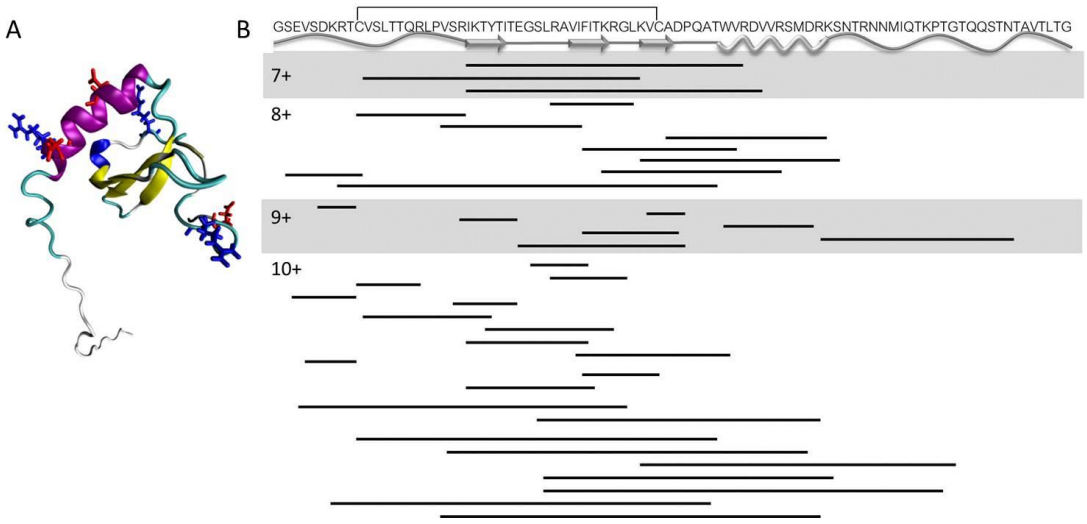

Figure 4.7: A) WT Ltn10 (PDB 1J9O) with residues implicated in salt bridge formation labelled acidic (red) and basic (blue), created in VMD. B) zIc fragmentation map for monomeric WT Ltn, with position of the disulfide bond (solid line) labelled.

For $[M_{WT}+7H]^{7+}$ and $[M_{WT}+8H]^{8+}$ little fragmentation is observed within the structural core, an exception to this is in the loop region between the end of the β-strand core and start of the α-helix (Ala52) where low intensity fragmentation is observed, which causes loss of residues 53-91 in a single z-type fragment, suggesting the α-helix is separable from the core. In solution the α-helix is positioned with respect to the β-sheets *via* a number of hydrophobic interactions[1]. Hydrophobic interactions can be lost in the gas-phase[10], which would explain separation of the α-helix from the β-sheet core. The salt bridges between Asp57-Arg60 and Arg60-Asp63 can provide intra-helix stabilisation enabling it to be lost in a single stable fragment and correspondingly, very little fragmentation along the α-helical region (residues 54-66) for $[M_{WT}+7H]^{7+}$ and $[M_{WT}+8H]^{8+}$ is observed. Loss of the α-helix is at a relatively low intensity and this could be due to the (lesser occupied) salt bridge between β1 and α-helix (Asp57-Arg17) which would limit α-helix dissociation and help to overcome the loss of hydrophobic interactions.

For the higher charge states ($[M_{WT}+9H]^{9+}$ and $[M_{WT}+10H]^{10+}$) significant fragmentation is observed along the α-helical region in addition to more extensive termini fragmentation. This suggests that the higher charge states are most likely due to a form of Ltn10 where unfolding in the C-terminus and partial unravelling of the α-helix has occurred, and could indicate a loss of the influence of Asp57-Arg17 and intra-helix (Asp57-Arg60 and Asp63-Arg60) interactions, allowing cleavage to be accompanied by dissociation. When considering $[M_{WT}+9H]^{9+}$ it is important to note that in DT IM-MS experiments two conformations were observed. The more extended conformation has a CCS of 1685 $Å^2$ and has been considered as a conformer far along the Ltn10 unfolding pathway. The high CCS (~10 % increase in

CCS between $[M_{WT}+8H]^{8+}$ and the most extended $[M_{WT}+9H]^{9+}$), in combination with the ECD fragmentation maps highlight that between charge states 8 and 9+ the α-helical region becomes unravelled and more amenable to cleavage. For $[M+10H]^{10+}$ fragmentation within the α-helical region and now along the ID tail indicates that loss of the α-helix disrupts the tails interaction with the core further accompanied by substantial reordering of the C-terminal region.

For all charge states, fragmentation within the β-sheet region is limited, which as discussed above is attributed in part to the disulfide bond. Considering the solution-phase structure, visual examination of the PDB file identifies a potential classical salt bridge comprising Glu30-Arg34 and Arg34-Asp49, positioned across the top of the triple stranded β-sheet. MD simulations for Ltn10 in water suggest these interactions can occur (Appendix 3 Figure A3.6A) at low occupancy (lower than those listed in Appendix 3 Table A3.4). *In vacuo* simulations demonstrate that this bridge would strengthen significantly within the gas-phase (Appendix 3 Figure A3.6B). If present this salt bridge could provide an intriguing explanation for the gas-phase stability of the β-sheet region, in combination with the disulfide bond.

For $[M_{WT}+8H]^{8+}$ a single, low yield, c-type fragment ion is observed at residue 68, the region between the α-helix and start of the tail, which results in cleavage of the ID tail. Fragmentation in this region increases in percentage yield and number of fragmentation sites as charge state increases, suggesting fragmentation of the ID tail becomes more favourable as the charge state increases, consistent with DT IM-MS experiments which attributed the unfolding event occurring between charge states seven and eight to tail unfolding. Furthermore, the MD simulations of $[M_{WT}+7H]^{7+}$ also provide evidence for the ID tail folding in towards the structural core in solution (Appendix 3 Figure A3.4), and the tail remaining associated in the gas-phase resulting in a compact CCS ~1250 $Å^2$ (Appendix 3 Figure A3.5) providing further support to the interpretation of the DT IM-MS data. Fragmentation in the region between the α-helix and ID tail is, however, of low yield with fragmentation occurring along the α-helix consistently observed at higher yield, highlighting that the α-helix is more amenable to cleavage than the region between the α-helix and ID tail in these studies.

Due to the low efficiency of ECD fragmentation, in order to carry out these studies a sufficiently high number of ions have to be trapped in the ICR cell prior to fragmentation. As all charge states of a protein are not observed at the same intensity and in order to trap a sufficient number of ions, the trapping times for all ECD experiments in this chapter were

tuned for each charge state, employing the minimum trapping time in all cases (all on the order of seconds). The possibility of ions unfolding further due to increased trapping time can therefore not be discounted and it is likely that the observed unfolding is due to a combination of both charge and trapping time in the ECD studies. In order to test this $[M_{WT}+10H]^{10+}$ was analysed after trapping for one and seven seconds (Appendix 3 Figure A3.7). The fragmentation maps obtained display similar trends however slightly more fragmentation in the β-sheet core is observed in the species trapped for seven seconds. This observation suggests that Ltn ions can undergo limited unfolding over the timescales of the ECD experiments. Unfolding as a function of time is an unavoidable consequence of such studies considering multiple charge states, as in addition to various other instrumental parameters the trapping time has to be tuned for each individual species, leading to increased trapping times for low intensity charge states. Matching the trapping times for all ions by increasing the time for higher intensity charge states would also have detrimental effects, namely oversaturation and signal ringing. The differences observed for increasing trapping time are much smaller, however, than the differences observed for increasing charge state, determined through comparisons of $[M_{WT}+8H]^{8+}$ and $[M_{WT}+9H]^{9+}$ (Figure 4.6) which were both trapped for the same length of time prior to ECD analysis but show significant differences in fragmentation. Therefore, in the remaining analysis of Ltn and its mutants unfolding as a function of trapping time is considered a minor contributor to this process and not discussed further. The major contribution to differences in ECD fragmentation maps and unfolding is attributed to charge state and hence the solution conformation prior to ionisation. In order to apply this technique to future protein systems, however, the effect of trapping time on unfolding for each protein would need to be assessed, as it may be expected to have a larger effect on less structurally stable proteins or proteins lacking any constriction due to disulfide bonds.

Comparison between the structural elements of Ltn10, gas-phase fragmentation maps and DT IM-MS bears up remarkably well for monomeric Ltn, and further careful consideration can yield additional information. The NMR studies on Ltn10 found that residues 1-10 have less NOE restraints and higher RMSD values[38], corresponding to a dynamic terminal region, and a region in which extensive backbone fragmentation was observed during ECD studies. In contrast residues 11-23 were found to be more ordered, with long-range NOEs and backbone RMSDs below 0.5 Å which has in part been attributed to the disulfide bond[38] and in ECD studies less fragmentation is observed in this region. Hence, both the core structure and different extents of disorder in both termini are distinguished in ECD fragmentation maps. Careful consideration of the fragmentation maps delineates the stability of each

structural region with respect to unfolding, highlighting that for Ltn10 the order of stability of each region is as follows C-terminus<N-terminus<α-helix<β3<β2<β1. Additionally, considering CCS in combination with the cleavage sites and fragmentation maps of the monomer over all charge states, insights into fold stability can be obtained (Figure 4.8).

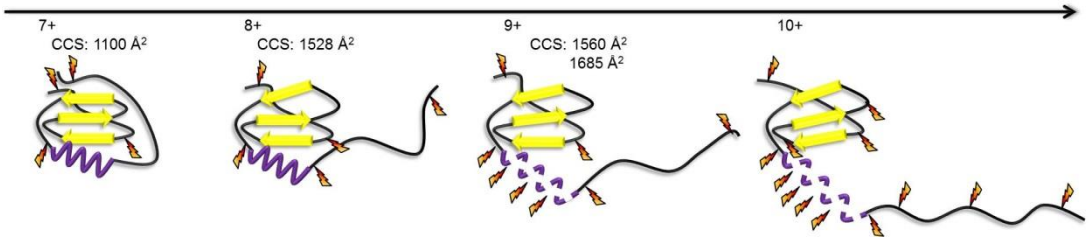

Figure 4.8: Cartoon representation of the conformers observed for Ltn10 over the charge states $[M_{WT}+7H]^{7+}$ to $[M_{WT}+10H]^{10+}$. Orange 'lightning bolts' indicate sites where ECD fragmentation is most significant.

4.3.2.4 ECD of monomeric Ltn WT 1-72

To further probe the fold and unfolding of Ltn10, monomeric WT 1-72 (M_{1-72}) was also subjected to ECD fragmentation studies. WT 1-72 can be thought of as the minimum model of structured fold for Ltn10. This construct is mostly monomeric, however, a small proportion of dimer is also observed (Appendix 3 Figure A3.2), which due to its low intensity is not discussed here.

For the lowest charge states of WT Ltn10 the ID tail is observed to be wrapped around or associated with the core, perhaps providing structural stability to the core, a hypothesis which can be investigated through ECD studies of WT 1-72. For WT 1-72 charge states $[M_{1-72}+5H]^{5+}$ to $[M_{1-72}+8H]^{8+}$, were observed at high enough intensity to be trapped and subjected to ECD fragmentation. The resulting zIc, c- and z-type fragment ions were identified and percentage yields calculated, shown in the fragmentation map below (Figure 4.9). For charge states $[M_{1-72}+5H]^{5+}$ to $[M_{1-72}+7H]^{7+}$ an increase in both the yield and extent of backbone fragmentation is observed with increasing charge states. For $[M_{1-72}+8H]^{8+}$ the extent of fragmentation occurring in the α-helix decreases in comparison to $[M_{1-72}+7H]^{7+}$, this is attributed to the lower signal intensity of the $[M_{1-72}+8H]^{8+}$ species, leading to fewer ions being trapped and subjected to ECD and highlights the difficulties of performing such ECD studies on multiple charge states. Despite the high sensitivity of the FT ICR-MS instruments, the low efficiency of ECD requires a signal intensity of greater than 1×10^6 ions for each scan, which can be overcome in many instances by increasing the trapping time allowing greater signal accumulation prior to ECD. For $[M_{1-72}+8H]^{8+}$ despite a trapping time of 10 seconds the signal intensity was on threshold, and lower than that of $[M_{1-72}+7H]^{7+}$, resulting in lower level backbone fragmentation being observed.

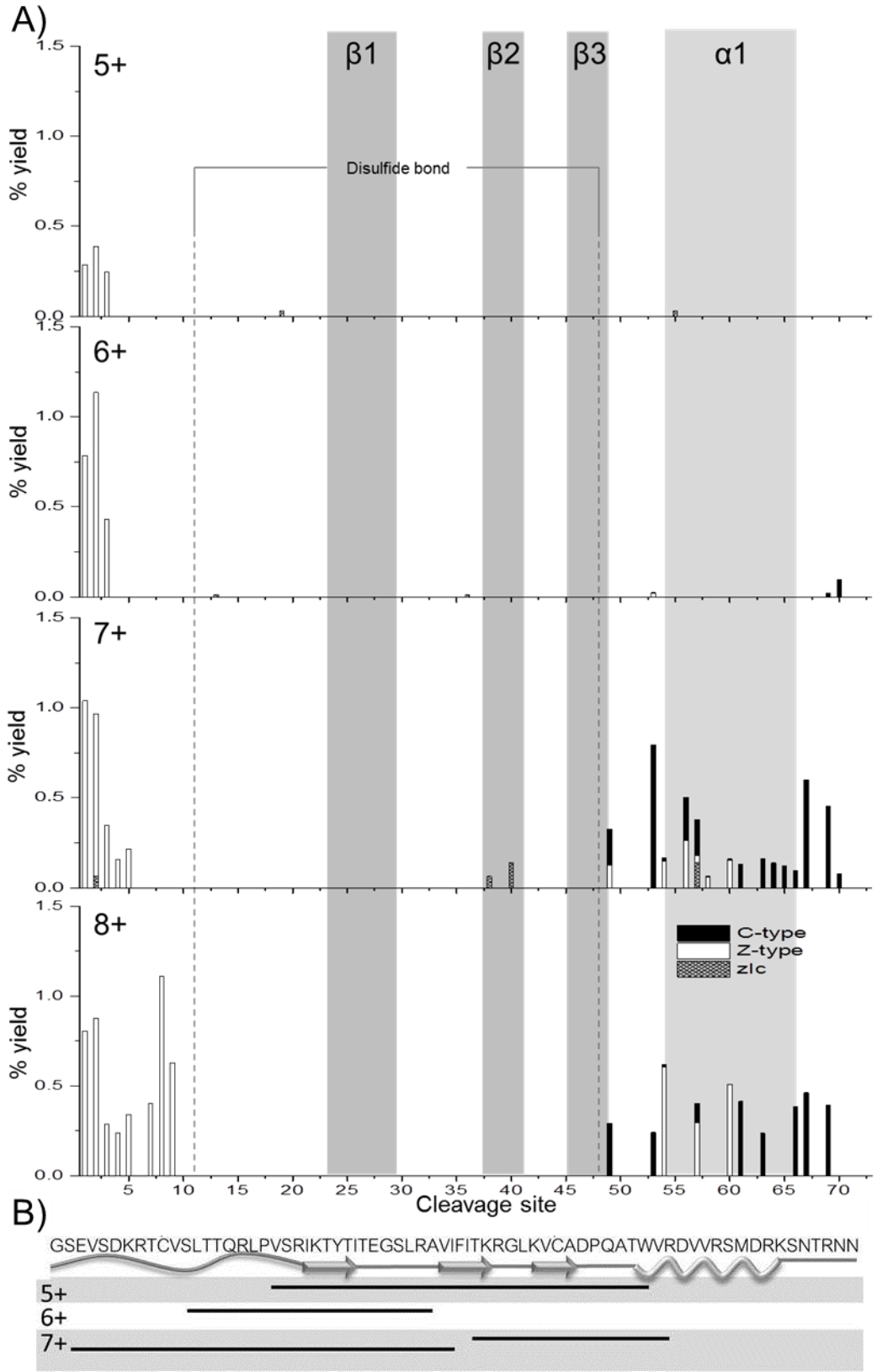

Figure 4.9: A) Percentage yields calculated for zIc, c- and z-type fragments as a function of cleavage site for $M_{1\text{-}72}$, charge states $[M_{1\text{-}72}+5H]^{5+}$ to $[M_{1\text{-}72}+8H]^{8+}$. B) Sequence of WT 1-72 labelled with disulfide and secondary structural elements, annotated with all zIc fragments observed.

Information on the fold and unfolding can be gleaned through the consideration of the fragmentation maps, as with WT Ltn10 fragmentation does not occur uniformly and can be mapped onto the secondary structural elements, enabling trends in unfolding to be discerned. As the charge state increases both the extent and intensity of fragmentation observed increases in the N-terminal region suggesting it is beginning to unfold from the structural core. In addition for $[M_{1-72}+7H]^{7+}$ and $[M_{1-72}+8H]^{8+}$ fragmentation in the α-helical region is observed suggesting this region is also beginning to unravel and is hence more amenable to fragmentation. It is interesting to note that two distinct conformations are observed for $[M_{1-72}+7H]^{7+}$, a compact conformation similar in CCS to those observed for $[M_{1-72}+4H]^{4+}$ to $[M_{1-72}+6H]^{6+}$ and a more extended conformation (>40 % larger than $[M_{1-72}+4H]^{4+}$). This more extended conformation is similar in CCS to that of $[M_{1-72}+8H]^{8+}$, which when combined with the results of ECD fragmentation, suggests that here the increase in CCS is due to unfolding of the α-helix.

As with full length WT Ltn10 very little fragmentation is observed in the β-sheet core, ascribed to greater stability in this region than the α-helix due to the disulfide bond and stability provided through gas-phase salt bridge formation, discussed above for full length WT Ltn. The fragmentation maps obtained from WT 1-72 and resulting trends are similar to those observed for WT Ltn10 with the protein unfolding *via* the C-terminus, and suggests that in the absence of the ID tail the fold is still stable and preserved in the gas-phase. One notable difference is that in full-length Ltn10 significant unravelling of the α-helix is not observed until $[M_{WT}+9H]^{9+}$ whereas in WT 1-72 this unravelling is observed from $[M_{1-72}+7H]^{7+}$. At lower charge states the ID tail is believed to be associated with the structural core of full-length WT protein and thought to unfold before the α-helix unravels. The fact that the α-helix is observed to unravel at lower charge states in WT 1-72 suggests that the ID tail may confer some form of structural stability and protection to WT Ltn10 fold and absence of the tail in WT 1-72 causes the α-helix to unfold at a lower charge states. Previous reports have shown that the ID tail of Ltn is entirely dispensable for receptor binding[38], questioning the importance of this region. If the tail is conferring stability or protection to the core it could help explain why this region is conserved. Alternatively, WT 1-72 may be observed to unfold at lower charge states due to its smaller size, as the charges are likely to occupy sites closer to one and other than in the full length protein at the same charge state, causing the protein to unfold to greater extents at lower charge states due to Coulombic repulsion. It is also possible that unfolding occurs due to a combination of both decreased stability due to the absence of the tail and greater Coulombic repulsion due to the smaller size of this construct. As with WT Ltn unfolding can be visualised through the

combination of DT IM-MS and ECD experiments highlighting vulnerable regions in the unfolding pathway (Figure 4.10).

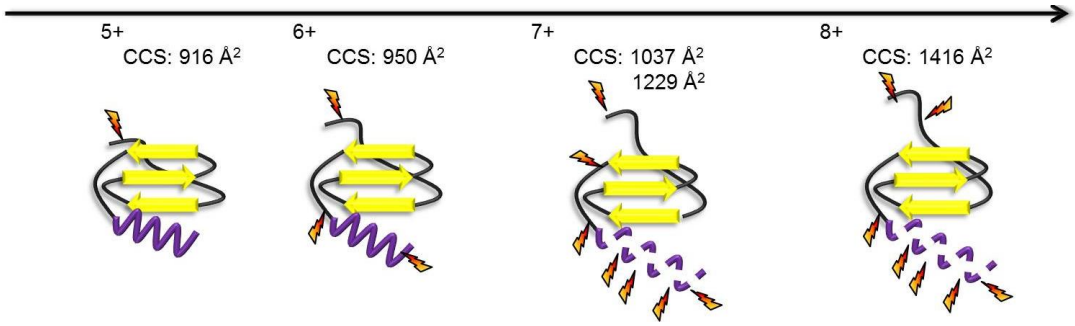

Figure 4.10: Cartoon representation of the conformers observed for WT_{1-72} Ltn10, for charge states $[M_{1-72}+5H]^{5+}$ to $[M_{1-72+}8H]^{8+}$. β-strands are represented by arrows and the α-helix is represented by a spiral, dashed spiral represents a loss of defined structure. Orange 'lightning bolts' indicate sites where ECD fragmentation is most significant.

4.3.2.5 ECD of dimeric WT Lymphotactin, Ltn40

To probe Ltn40, only the charge states of dimer where net charge is an odd number and therefore are not coincident with monomer species were studied, namely $[D_{WT}+9H]^{9+}$, $[D_{WT}+11H]^{11+}$ and $[D_{WT}+13H]^{13+}$. As for the monomer, zIc, c- and z-type fragments were considered; however, since both monomeric units are identical in sequence these experiments do not distinguish between the two chains and therefore the total fragment percentage yields are mapped onto a single monomeric unit (Figure 4.11 and Figure 4.12). There are two main fragmentation pathways for this non-covalent complex; dimer can remain intact and fragment, or dimer can dissociate to monomer and then fragment. Both were considered during analysis, shown in Figure 4.11 where above axis histograms represent dimer fragmentation and below axis chromatograms represent dissociation to monomer followed by fragmentation.

For $[D_{WT}+9H]^{9+}$ ECD mainly leads to dissociation from dimer to monomer followed by extensive fragmentation along the backbone, which suggests that the dimer interface is not stable in this form. The instability of the dimer at this charge state can be attributed to either collapse upon ionisation and transfer or to a less stable encounter species, and is entirely consistent with the findings from DT IM-MS, in which the CCS for the lowest charge states of dimeric WT Ltn are smaller than that of the predicted CCS, a phenomena which has been reported previously[11,48].

The two higher charge states studied by ECD ($[D_{WT}+11H]^{11+}$ and $[D_{WT}+13H]^{13+}$) show greater stability as dimers, with a much lower proportion of dimer dissociating to monomer and then fragmenting, instead retaining the dimer interface and fragmenting. The resulting

fragmentation maps are striking, showing almost complete retention of the dimer within the β-sheet interface, residues 11-51. The majority of fragments arise from cleavage either side of this structural core, residues 1-10 and 47-61 indicating that this β-sheet region remains stable within the gas-phase. The results suggest that the ID tails are wrapped around the structural core, and that the less dynamic part of this tail (residues 52-60) is more closely associated.

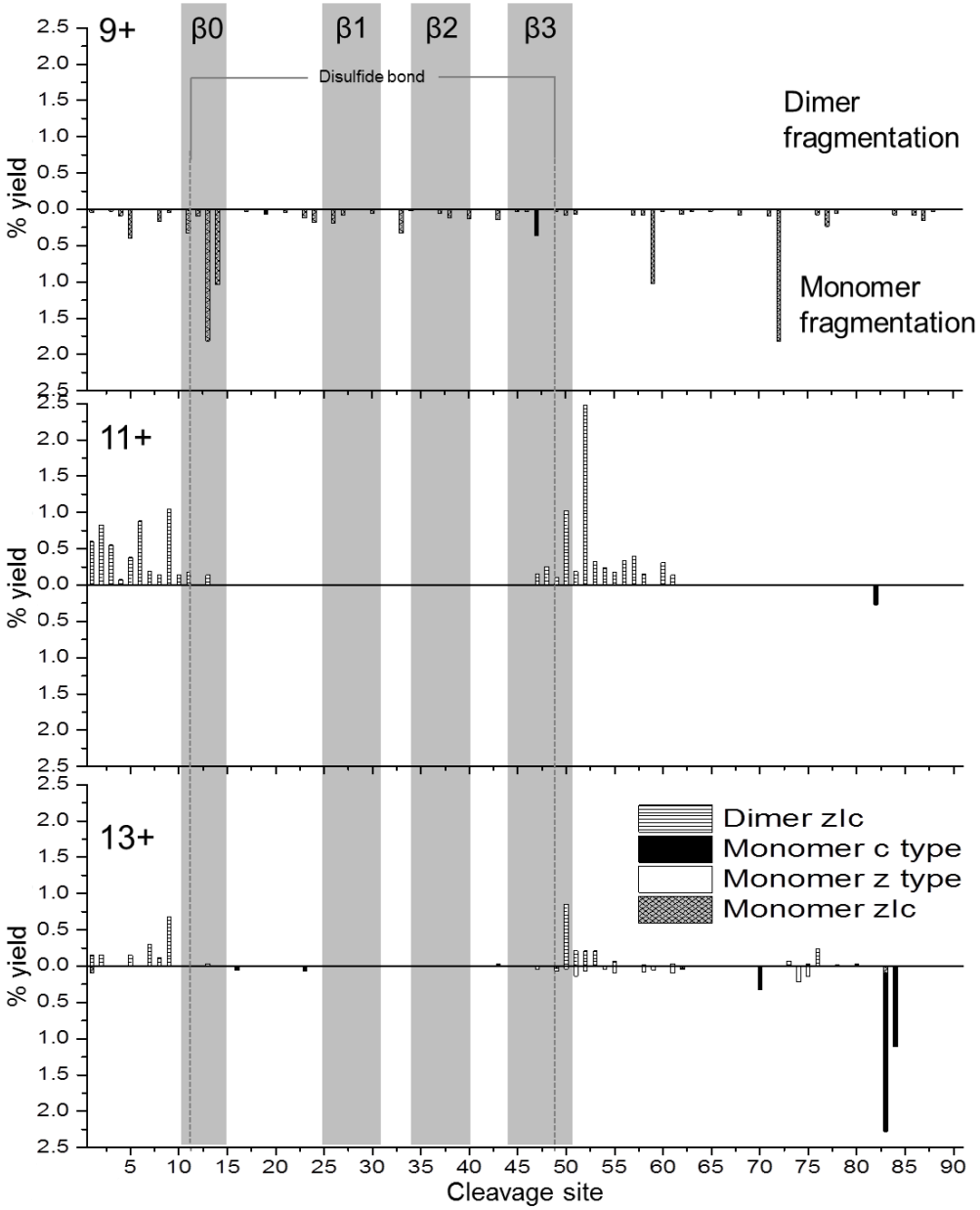

Figure 4.11: Percentage yield calculated for zIc, c-and z-type fragments as a function of cleavage site for dimeric Ltn $[D_{WT}+9H]^{9+}$, $[D_{WT}+11H]^{11+}$ and $[D_{WT}+13H]^{13+}$. Above axis histograms represent fragments formed while retaining the dimer interface. Below axis histograms represent fragments formed following dimer dissociation to monomer.

As with Ltn10 possible interactions which could serve to stabilise Ltn40 within the gas-phase were considered, in order to better understand the stability of $[D+11H]^{11+}$ and $[D+13H]^{13+}$. MD simulations were performed in solution and *in vacuo* to identify any potential salt bridges, which would act to stabilise the gas-phase structure. A number of salt bridges were identified in the two solution runs (Appendix 3 Table A3.6) and the *in vacuo* run (Appendix 3 Table A3.7). As with Ltn10 the occupancy of some identified salt bridges increased for Ltn40 *in vacuo* in comparison with in solution, characteristic of electrostatic interactions, and analysis will focus on salt bridges retained in the gas-phase. *In vacuo* MD simulations identified a number of intra-chain salt bridges between residues Asp49-Arg8, Asp49-Arg34 for Chain A and interactions between Asp6-Lys7, Asp49-Arg8, Asp49-Arg34 and Asp57-Arg56 for Chain B. A higher number of salt bridges are identified in Chain B due to the asymmetric distribution of charges in the 11+ species, where Chain A contains an additional charge. All identified intra-chain salt bridges lie on the edges of the β-sheet core and hence if these interactions are preserved in the gas-phase, they would stabilise this structural region. Furthermore, in Ltn40 the disulfide bond covalently links β0 to β3 providing further intra-chain stability.

A high number of possible inter-chain salt bridges were also identified; Arg22-Glu30, Glu30-Lys24, Arg22-Asp57, Asp57-Arg17, Arg42-Glu3, Arg17-Glu3, Glu30-Arg61, and Glu30-Arg22 (denoted ChainA-ChainB). Significantly all of these potential salt bridges lie on the edges of the β-strands and could therefore strengthen the dimer interface (Figure 4.12). These stabilising interactions could serve to retain aspects of the Ltn40 fold in the gas-phase, which is supported by the finding that these residues are retained at high frequency in zIc fragments (Figure 4.12).

Figure 4.12: A) Ltn40 (PDB 2JP1) with inter-chain salt bridges labelled, acidic residues (red) and basic (blue), created in VMD. B) WT sequence with disulfide bond (solid line). All stretches of internal fragments observed in one or more zIc fragments for the $[D+11H]^{11+}$ and $[D+13H]^{13+}$ species, the summed intensity of internal fragments with respect to cleavage site is shown in Figure 4.11.

The CCS of $[D+11H]^{11+}$ as a function of time during the 200 ns MD simulation is centred at ~1550 Å^2, smaller than the experimental value (1802 Å^2) for this species as determined by DT IM-MS (Appendix 3 Figure A3.8). Differences here can be attributed to the C-terminal tail which is not present in the starting structure (PDB 2JP1) and would certainly increase the CCS. Careful comparison of the results from the combination of gas-phase techniques presented enables the unfolding pathway of this protein to be mapped (Figure 4.13).

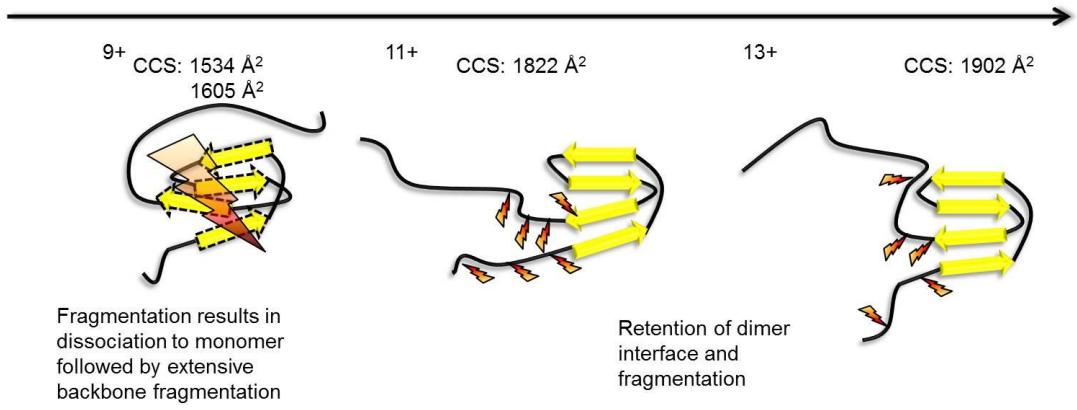

Figure 4.13: A cartoon representation of the conformers observed for Ltn40 for charge states $[D+9H]^{9+}$, $[D+11H]^{11+}$and $[D+13H]^{13+}$. For $[D+11H]^{11+}$and $[D+13H]^{13+}$ the dimer is represented here by a single monomeric unit with β-strands being represented by arrows. Orange 'lightning bolts' indicate sites where ECD fragmentation is most significant for the dimer. For $[D+9H]^{9+}$ where fragments only map to a monomer unit with no secondary structure, the extensive fragmentation is depicted by a large lightning bolt and dashed lines represent a lack of defined structural elements.

In order to determine if dimeric Ltn is present in the Ltn40 conformation as opposed to a less specific form such as a dimer of Ltn10, comparisons between the two fragmentation maps are made (Figure 4.6 and Figure 4.11). The residues 10-25 in Ltn10 have little structure and hence fragmentation is observed along the backbone here. In Ltn40, however, these residues have formed the β0 strand and loop to β1, and subsequently in dimeric WT Ltn very little fragmentation is observed in this region, consistent with increased structural stability. Next considering the residues which form the α-helical region in Ltn10 (residues 54-66), ECD fragmentation maps of the higher charge states of monomeric Ltn show sequencing along the entire backbone of this region, consistent with an α-helical region unravelling. In dimeric Ltn, fragmentation is not observed along this region and instead fragmentation is observed to branch out from the β-sheet core, with the most favourable fragmentation site observed as residue 53, demonstrating that retention of the structural core and loss of residues 52-93, which contain very little secondary structure in Ltn40, is preferential. Additional fragmentation in the nine amino acids which form the first part of the ID tail (residues 52-60) is observed in Ltn40 suggesting this region is associated with the protein core and hence fragmentation along this region is possible. This is comparable with the findings from NMR

which allowed this region to be solved. Fragmentation maps of monomeric and dimeric Ltn show prominent differences between the two, providing compelling evidence for retention of both the Ltn10 and Ltn40 conformations in the absence of solvent, and the power of this technique to report on folded forms and unfolding intermediates. These significant differences in the ECD fragmentation of the two WT Ltn species, highlight that here it is probable that both distinct folds of this metamorphic protein have been preserved and can be differentiated in an unsolvated form.

4.3.3 Probing the Ltn10 and Ltn40 folds through specific site mutations

The Volkman research group previously designed two specific site mutants CC3 and W55D, intended to mimic each distinct fold of Ltn whilst limiting interconversion to the other[3,38]. CC3 and W55D are studied here to further probe the Ltn10 and Ltn40 folds in the gas-phase and are compared to WT Ltn. All experimental conditions for both CC3 and W55D, were kept as similar to WT Ltn as possible to enable direct comparisons to be made, enabling each distinct fold to be further probed and the effectiveness of these mutations to stabilise each fold to be determined.

4.3.3.1 Probing the Ltn10 fold, consideration of the CC3 mutant

CC3 is designed to mimic Ltn10 and contains an additional disulfide bond based on the third disulfide in the CC chemokine, HCC-2[38]. CC3 was engineered by introducing two extra cysteines in positions 21 and 59, with the aim that addition of an extra disulfide bond, based on a chemokine fold, would structurally constrain Ltn stopping structural rearrangement to Ltn40. The spectrum obtained for CC3, highlights that this mutant exists primarily in a monomeric (M_{CC3}) form with very little dimer signal as expected (Figure 4.14A). The CC3 monomer presents over a range of charge states from $[M_{CC3}+5H]^{5+}$ to $[M_{CC3}+8H]^{8+}$, with species which accept either six or seven protons being the most favourable and hence most intense.

The global fold of CC3 was assessed by DT IM-MS and the CCS determined are consistent with a structurally constrained species presenting over a narrow range of conformations (Figure 4.14B and Appendix 3 Table A3.8) increasing by only 23 %, a stark contrast to WT Ltn which was observed to increase by up to 50 %. This decrease in flexibility is expected considering CC3 contains a second disulfide bond. The NMR structure for CC3 has been solved (PDB 2HDM) enabling direct comparisons between experimental and theoretical values for this construct to be made, and are found to be in extremely good agreement. The file used to determine the theoretical CCS however does not contain the final 18 amino acids

in the ID tail. As experimental values for the full length protein match extremely well with this theoretical value it is surmised that, as with WT Ltn, the ID tail is wrapped around or associated with the structural core. In the case of CC3, however, the ID tail is associated with the core at all charge states studied, attributed in part to the disulfide bond placement which pins the α-helix to the structural core and may therefore also encourage tail association increasing the stability of this compact form and limiting unfolding.

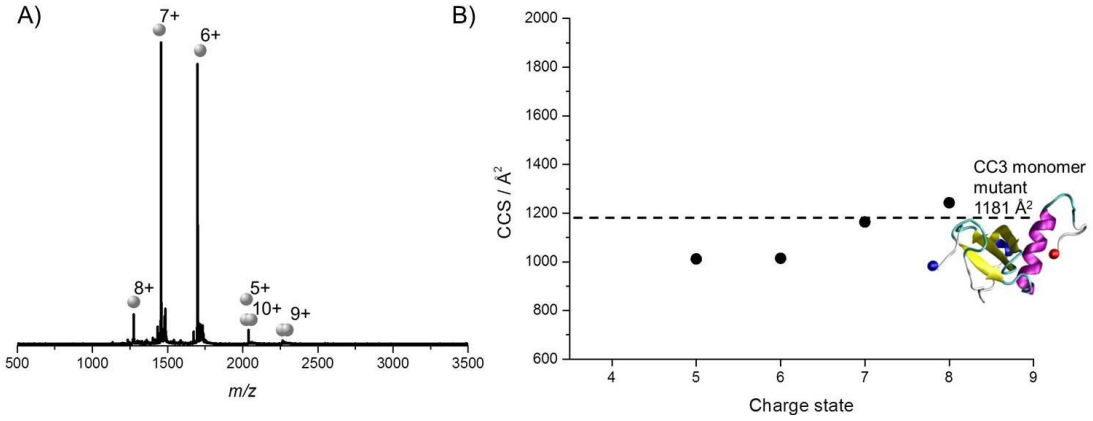

Figure 4.14: CC3 Ltn in 20 mM AmAc A) Mass spectra obtained for 50 μM protein B) Experimental CCS, reported values are an average of three different day repeats. Error bars fall within the symbol size and represent standard deviation of the three repeats. Theoretical CCS of the CC3 mutant was determined from the NMR structure (PDB 2HDM) and is shown by dashed line.

To probe the stabilising effect of the additional disulfide bond on the Ltn10 fold and unfolding pathways, ECD fragmentation studies were performed on CC3. $[M_{CC3}+5H]^{5+}$ to $[M_{CC3}+7H]^{7+}$ were observed at high enough intensity to be subjected to ECD fragmentation and the resulting fragmentation maps are shown in Figure 4.15. For all species studied very little fragmentation is observed, as expected due to their relatively low charge states and compact nature. For $[M_{CC3}+5H]^{5+}$ and $[M_{CC3}+6H]^{6+}$ only minor fragmentation is observed in both termini, with a marginal increase in backbone fragmentation observed for the two remaining charge states ($[M_{CC3}+7H]^{7+}$ and $[M_{CC3}+8H]^{8+}$). The fragmentation maps highlight, at all charge states, CC3 is present in compact stable conformations which cannot undergo extensive fragmentation, an observation which is entirely consistent with findings from DT IM-MS. It is important to note that fragmentation would also be limited due to the two disulfide bonds, which could be preferentially cleaved in ECD resulting in a lower yield for backbone fragmentation[46].

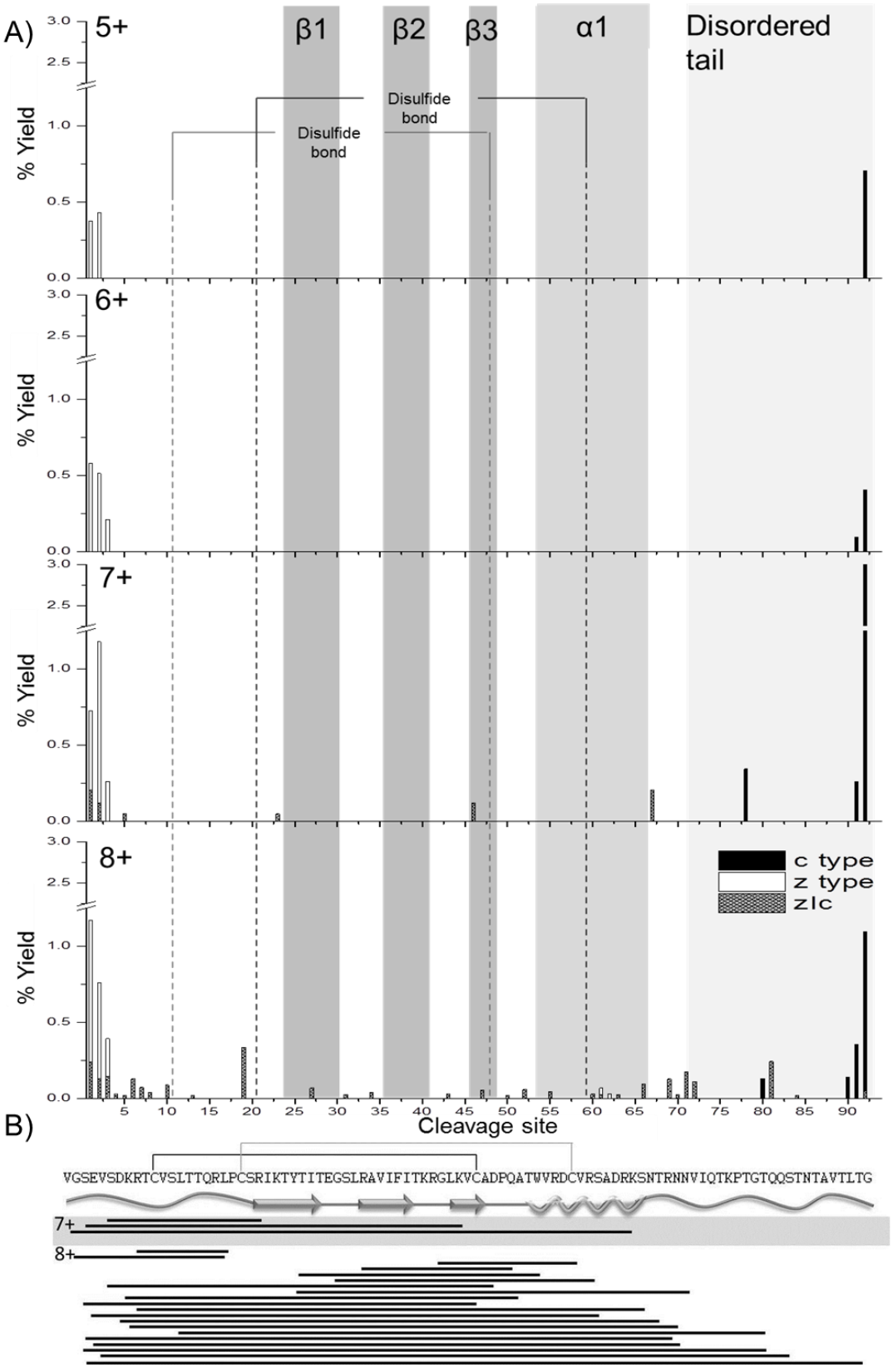

Figure 4.15: A) ECD fragmentation maps over charge state range $[M_{CC3}+5H]^{5+}$ to $[M_{CC3}+8H]^{8+}$, species formed are identified as zIc, c- and z-type fragments. B) CC3 sequence labelled with disulfide bond (solid line) and position of all zIc fragments observed for CC3 under buffered conditions.

For both $[M_{CC3}+7H]^{7+}$ and $[M_{CC3}+8H]^{8+}$, zIc fragments corresponding to complete loss of the ID tail are observed, suggesting at these higher charge states the tail is more loosely associated with the structural core and hence can be lost during ECD fragmentation. For $[M_{CC3}+8H]^{8+}$ low intensity fragmentation in the α-helix is also observed suggesting this is the first secondary structural element to unfold, consistent with WT Ltn10. Complete sequencing along the backbone of this region is however not observed suggesting it has not unravelled and unfolded fully an observation consistent with DT IM-MS experiments in which $[M_{CC3}+8H]^{8+}$ is only 23 % larger than the most compact species.

As with WT Ltn, minimal fragmentation occurs within the β-sheet core, with fragmentation in this region forming low intensity zIc fragments (Figure 4.15); 19 % of the total fragments for this species are due to fragmentation here, corresponding to only 6 % of the total intensity. Indeed considering the stabilising interactions present it is anticipated that very little fragmentation would occur within the core, as CC3 contains two disulfide bonds between residues 11-48 and residues 21-59, which would act to stabilise this core region substantially. Furthermore, as with Ltn10 there are likely to be a number of stabilising salt bridges conferring stability to the structure, within the α-helix and perhaps due to the compact nature of these species within the structural core. In-depth salt bridge analysis was, however, not performed here.

Considering DT IM-MS and ECD data, and the fact that this mutant is mostly observed at low charge states, it is clear that this construct is more compact and less prone to tail unfolding than WT Ltn and hence CC3 is an effective, stable construct to mimic the Ltn10 fold.

4.3.3.2 Probing the unfolding landscape of Ltn10: supercharged CC3

To probe the fold and stability of CC3 at higher charge states and potentially further along the unfolding pathway CC3 was studied under similar instrumental conditions but after the addition of a 'supercharging' reagent, *meta*-nitrobenzyl alcohol (*m*-NBA). *m*-NBA causes an increase in charging due to an increase of surface tension in the *n*-ESI droplets[49], and due to charge repulsion the increase in charge can promote more extended species. The addition of 1 % *m*-NBA to 20 mM ammonium acetate enabled the highest observed charge state in the spectra to increase from $[M_{CC3}+8H]^{8+}$ to $[M_{CC3}+13H]^{13+}$ (Figure 4.16A). The CCS for all species were determined by DT IM-MS (Figure 4.16B and Appendix 3 Table A3.8). For charge states observed in both buffered and supercharged conditions, an increase in CCS of up to 16 % is detected in the supercharged conditions; suggesting *m*-NBA is promoting more

extended structures at all charge states. This is most evident for the highest charged species observed under both solution conditions, which show the greatest increase in CCS in the presence of *m*-NBA. The difference in CCS after addition of *m*-NBA suggests that the conformations are somewhat altered upon addition of this supercharging reagent. Indeed previous studies have shown that addition of supercharging reagents can disrupt the native conformation of protein complexes[50]. The promotion of higher charge states, however, enables the unfolding of CC3 to be further explored, which could not be afforded through consideration of only buffered conditions.

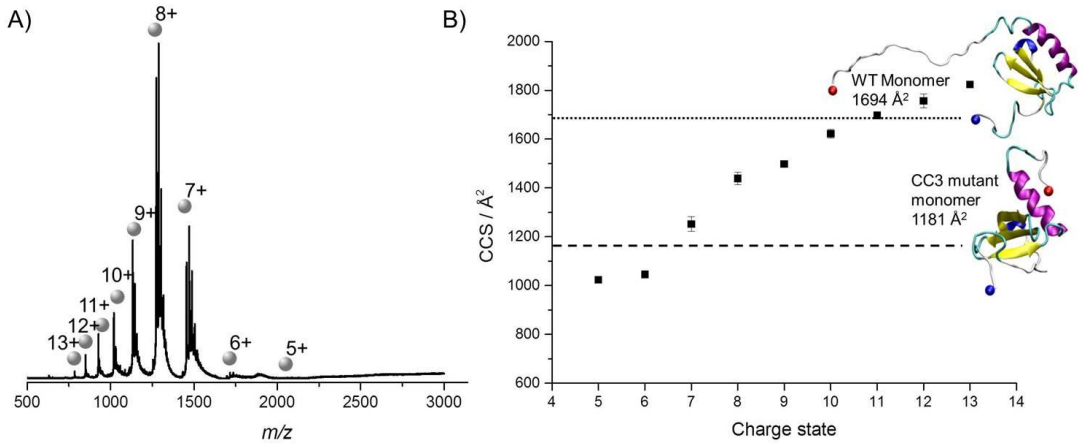

Figure 4.16: CC3 Ltn in 20 mM AmAc plus 1 % *m*-NBA A) Mass spectra obtained for 50 µM protein B) Experimental CCS, reported values are an average of three different day repeats. Error bars fall within the symbol size and represent standard deviation of the three repeats. Theoretical CCS of CC3 and WT Ltn were determined from the NMR structures (PDB 2HDM and 1J9O) and are shown by dashed and dotted lines respectively.

From Figure 4.16B it is evident that under these supercharged conditions, there is no major conformational shift or key unfolding transitional state, characterised by a significant shift in CCS observed at a single charge state. For WT Ltn a major conformational switch was observed between charge states 7+ and 8+, which was attributed to α-helix unravelling attributed to a loss of salt bridge stabilisation, however, for CC3 due to the placement of the additional disulfide bond (Figure 4.17) this unfolding pathway is limited. Instead CCS increases steadily (~1023-1824 $Å^2$) highlighting that under these conditions unfolding of CC3 is a continuous process consisting of small incremental increases in CCS. CCS obtained for supercharged CC3 Ltn are compared both to the theoretical CCS obtained from the NMR coordinates for this mutant (PDB 2HDM) and those obtained for WT Ltn10 (PDB 1J9O), with the ID tail extended out from the structural core of the protein (Figure 4.16B). The theoretical CCS determined for CC3 compares well with the lowest charge states observed for CC3 under these conditions ($[M_{CC3}+5H]^{5+}$ and $[M_{CC3}+6H]^{6+}$), suggesting that despite

addition of *m*-NBA they are most likely still folded with the ID tail associated with the structural core. Whereas, the theoretical CCS obtained from WT Ltn10 compares well with the experimental CCS for $[M_{CC3}+12H]^{12+}$ and $[M_{CC3}+13H]^{13+}$. As the α-helix in CC3 can be thought of as being stapled to the core and as comparisons between the highest charge states of CC3 bear up remarkably well with WT Ltn, the increase in experimental CCS is attributed to tail unfolding.

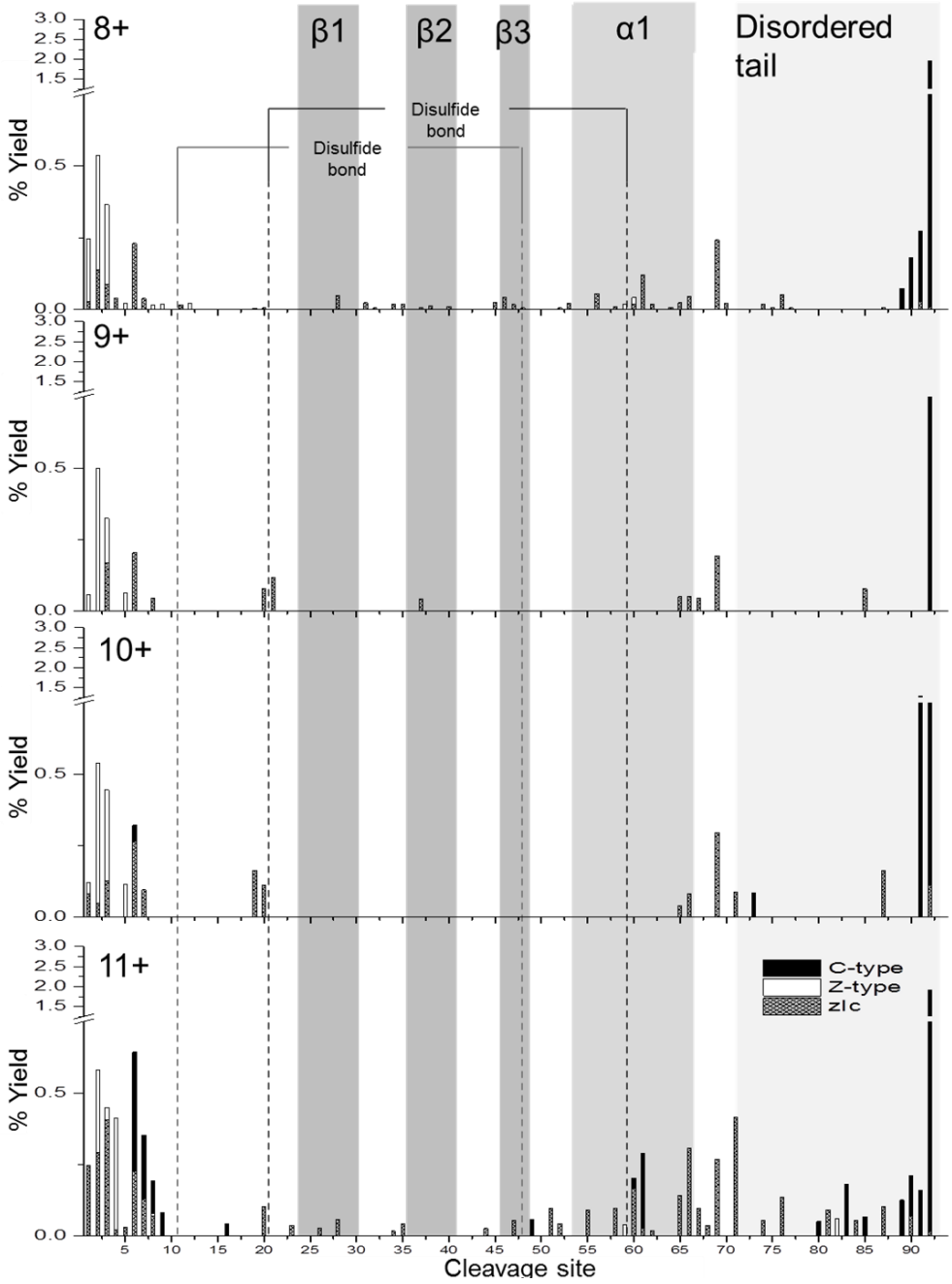

Figure 4.17: ECD fragmentation maps for supercharged CC3 over charge state range $[M_{CC3}+8H]^{8+}$ to $[M_{CC3}+11H]^{11+}$, species formed are identified as zIc, c-and z-type fragments.

To further examine the *m*-NBA promoted unfolding pathway, $[M_{CC3}+8H]^{8+}$ to $[M_{CC3}+11H]^{11+}$ were subjected to ECD fragmentation (Figure 4.17). It is interesting to first consider the $[M_{CC3}+8H]^{8+}$ species, which was studied under both buffered and supercharged conditions. Under supercharged conditions more extensive fragmentation is observed along the backbone particularly in the α-helical region, suggesting this region is more unfolded under these solution conditions, consistent with the observed increase in CCS which suggested unfolding was promoted in the presence of *m*-NBA.

Figure 4.18: CC3 sequence labelled with disulfide bond (solid line) and secondary structural elements, annotated with all zIc fragments observed for CC3 under supercharged conditions.

As charge state increases more extensive fragmentation is observed in both the α-helix and along the ID tail, emphasising that these regions are the first to unfold from the core of the protein. At all charge states studied under supercharged conditions zIc fragments are observed which correspond to the complete loss of the tail and regions of the N-termini, retaining the structural core (Figure 4.18). Thus under these conditions, the disordered regions are more weakly associated with the structural core and hence are more available for fragmentation. Additionally, for $[M_{CC3}+11H]^{11+}$ greater fragmentation in the α-helix is observed highlighting this region is becoming more unfolded and beginning to unravel from the structural core. An increase in fragmentation along the C-terminus is also observed for this charge state suggesting it has undergone a reordering event. For $[M_{CC3}+11H]^{11+}$ fragmentation within the β-sheet region is observed, at lower intensity than that observed for the α-helix and consistent with a compact, perhaps structured region which is stabilised by multiple interactions including the two disulfide bonds. As with WT Ltn when fragmentation

occurs in the β-sheet region it is mostly in the form of zIc fragments (Figure 4.17) thought to retain at least some of the stabilising interactions present in solution.

It is important to note that due to the conformational constriction implemented in this mutant by the addition of the second disulfide bridge, the unfolding pathway of WT Ltn10 involving complete loss of the α-helix and disordered tail in a single fragment is not observed until the most extended species ($[M_{CC3}+11H]^{11+}$) and is facilitated through ECD disulfide reduction, which emphasises that the additional disulfide structurally constrains the Ltn10 fold inhibiting the unfolding as expected. Combining ECD and DT IM-MS results for buffered and supercharged CC3 Ltn a more extensive picture of the unfolding of this mutant can be afforded, than that allowed by studying this protein only in buffered conditions. Even in the most extended species observed there is evidence that the structural core remains relatively intact and the observed fragments map extremely well to the Ltn10 fold, stressing that the fold and secondary structural elements comprising it are not significantly perturbed through addition of *m*-NBA. Addition of *m*-NBA clearly can promote extended structures, however, this could still be a useful reagent for use in the study of protein fold, particularly for proteins containing multiple disulfide bonds such as CC3.

4.3.3.3 Probing the Ltn40 fold through specific site mutation

The final mutant studied is a single point mutation (W55D), designed to push the equilibrium towards the dimeric Ltn40. The amino acid replacement is located in the beginning of the α-helix in the Ltn10 conformation and is designed to destabilise this conformation, with loss of the tryptophan side chain disrupting the Ltn10 hydrophobic core. In order to determine to what extent the W55D mutant mimics the Ltn40 fold, W55D was prepared under the same solution conditions (20 mM AmAc) and studied under the same instrumental conditions applied for all other Ltn species. Under these conditions the most intense species is monomeric (M_{W55D}) presenting over a range of charge states from $[M_{W55D}+4H]^{4+}$ to $[M_{W55D}+9H]^{9+}$ (Figure 4.19A). The dimeric species (D_{W55D}) is also observed, at somewhat lower intensity than the monomer, over the charge state range $[D_{W55D}+9H]^{9+}$ to $[D_{W55D}+13H]^{13+}$ (Figure 4.19A), wider by two charge states than WT Ltn dimer.

In order to determine if the equilibrium could be shifted towards the dimeric form a number of solution conditions were studied, including increasing buffer strength and alteration of the pH (Appendix 3 Figure A3.9), however, no significant increase in dimer was observed. The instrumental parameters were tuned to ensure gentle desolvation conditions, however,

monomer prevailed as the most intense species under all conditions, a significant point which is discussed in more detail below.

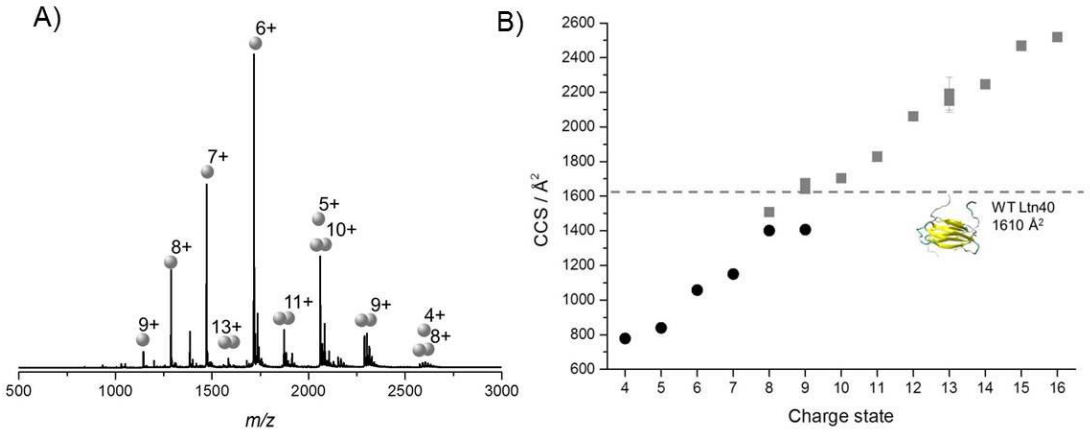

Figure 4.19: W55D Ltn. A) Mass spectra acquired in 20 mM AmAc pH 6.8 B) Average experimental CCS obtained from three different day repeats, error bars are calculated from the standard deviation and lie within symbol size. A wider charge state range of dimer is observed in B due to differences in the instrumentation used. Dashed line represents theoretical CCS of WT Ltn40 obtained from the NMR structure (PDB 2JP1).

From DT IM-MS it is clear that the monomeric global unfolding signature is different for W55D than for all other monomeric forms of Ltn studied previously under buffered conditions, with CCS increasing almost uniformly with charge from ~778 to 1407 $Å^2$ (Figure 4.19B and Appendix 3 Table A3.9). There are a number of plausible conformations that M_{W55D} could exist in, including the Ltn10 conformation, a single unit of Ltn40 or a yet unknown conformation. Consideration of CCS alone cannot provide an answer to this problem, as experimental CCS are observed to lie over a wide range encompassing both the theoretical CCS of a single Ltn40 chain without the ID tail (1141 $Å^2$) and that determined for Ltn 1-72 (1228 $Å^2$), which is thought to represent the Ltn10 fold with the tail wrapped around or associated with the structural core. Therefore, these species present an interesting challenge in identifying and studying protein fold; can this combination of DT IM-MS and ECD be used to determine which fold, if either, M_{W55D} exists in?

Hence $[M_{W55D}+6H]^{6+}$ to $[M_{W55D}+10H]^{10+}$ were subjected to ECD and observed fragment ions were subsequently mapped onto both folds (Figure 4.20). As expected, fragmentation increases as a function of charge, however, this does not occur uniformly and therefore trends observed reveal conformational information. The most extended, highest charged species ($[M_{W55D}+10H]^{10+}$), displays extensive fragmentation into the core residues of both folds suggesting that this species is far along the unfolding pathway resulting in many possible sites of fragmentation. For consideration of W55D monomeric conformations it is

therefore more appropriate to consider the intermediate charge states, in which the protein ions are beginning to unfold but are not yet highly extended and unfolded.

Beginning with $[M_{W55D}+8H]^{8+}$ fragmentation is observed in residues 52-65. This fragmentation pattern does not appear to be consistent with unravelling and unfolding of an α-helix present in the Ltn10 form as has been observed for WT Ltn10, which is characterised by cleavage along the backbone in this region (Figure 4.20A and Figure 4.21). Instead starting from $[M_{W55D}+8H]^{8+}$ fragmentation is observed to branch out from the core residues comprising the β-sheet core in the Ltn40 fold, with low intensity fragments also being observed within the core (Figure 4.20B and Figure 4.21). This observation is consistent with that of the WT Ltn in the Ltn40 fold, in which fragmentation is observed to branch out from the β-sheet core with higher intensity fragments in the first nine amino acids after the β3 strand, consistent with the NMR structure which shows that these regions are associated with the protein core allowing them to be solved. Indeed at all charge states, fragmentation is observed to branch out from the structural core of Ltn40 with minor fragmentation observed between strands suggesting M_{W55D} is perhaps existing as a single chain of Ltn40.

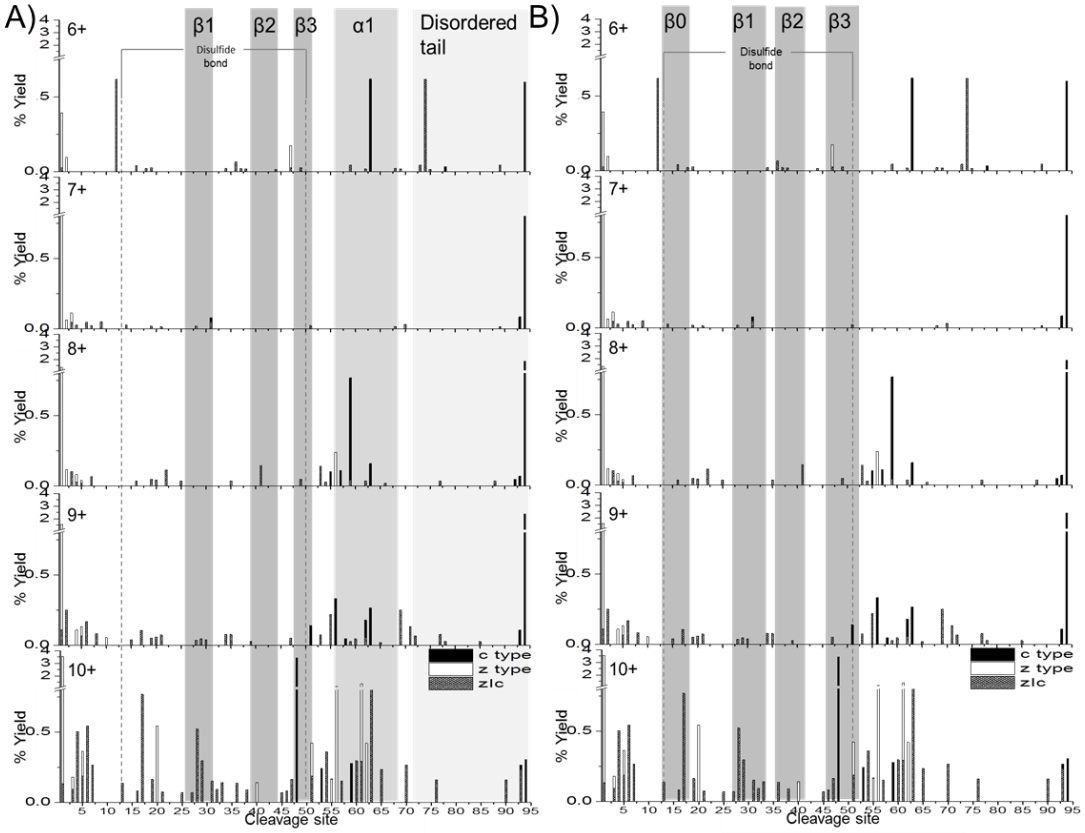

Figure 4.20: Monomeric W55D ECD analysis: Percentage yields calculated for c type, z type and zIc fragments as a function of cleavage site for M_{W55D} over the charge states $[M_{W55D+}6H]^{6+}$ to $[M_{W55D}+10H]^{10+}$. Fragments are mapped onto A) Ltn10 fold and B) Ltn40 fold. A y-axis scale break is included in all cases for figure clarity.

Furthermore, the distinct differences in DT IM-MS and ECD studies of this monomeric system in comparison to WT, WT 1-72 and CC3 Ltn, all of which appear extremely similar suggest these species are not in the Ltn10 fold. Similarities between the fragmentation maps obtained here with those previously obtained for Ltn40 (Figure 4.11) in combination with the NMR studies (which show a primarily dimeric species in the Ltn40 conformation) suggest it is conceivable that M_{W55D} could be a single chain of Ltn40, however, it is impossible to definitively determine which fold M_{W55D} is existing in from these studies. A monomeric form of Ltn40 could be a result of dissociation of the dimer upon the ionisation and transfer of the protein, suggesting that under our experimental conditions the dimer interface is not as stable as the WT Ltn40 interface. Alternatively a proportion of W55D could exist in a monomeric form in solution under these conditions. Both of these possible explanations suggest that this mutant is not a stable, constrained mimic of Ltn40 advising that the design of a more stable dimeric mutant is essential.

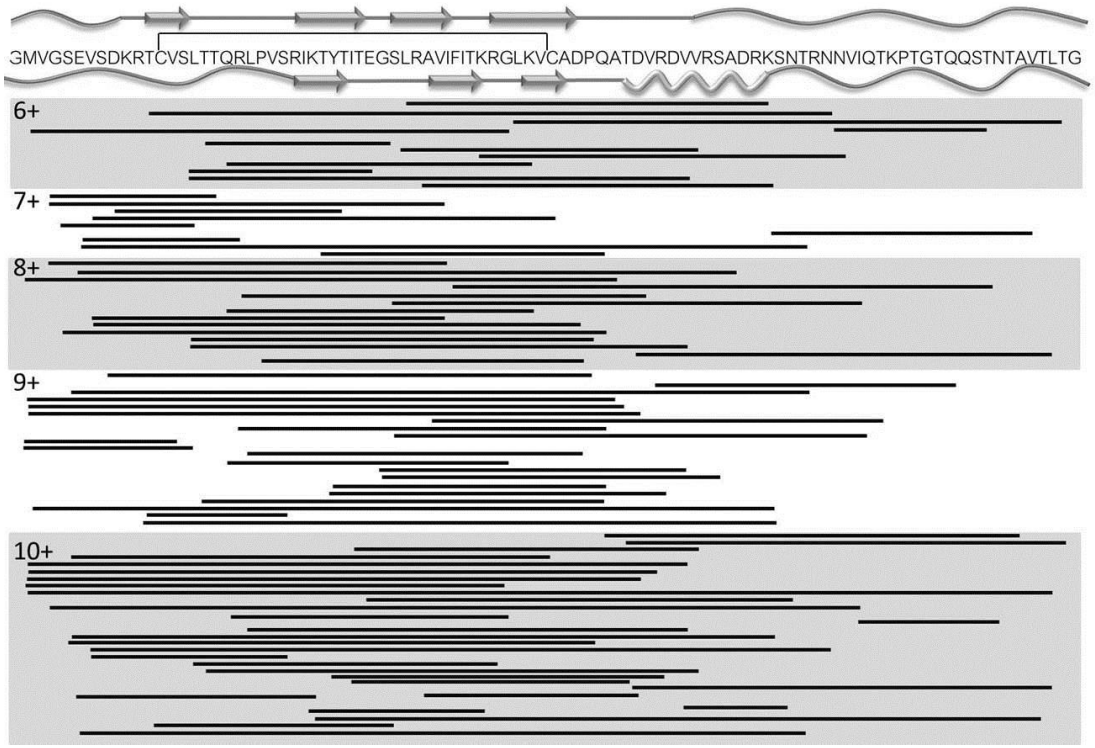

Figure 4.21: W55D sequence labelled with disulfide bond (solid line), the secondary structural elements of Ltn40 (above) and Ltn10 (below) folds, annotated with all zIc fragments observed for monomeric W55D.

DT IM-MS highlights dimeric W55D (D_{W55D}) (Figure 4.16B and Appendix 3 Table A3.9) is considerably different to D_{WT} (Figure 4.4). Dimeric W55D Ltn presents over a CCS range of ~1509 to 2519 $Å^2$, with the most extended form ($[D_{W55D}+16H]^{16+}$) being 32 % larger than the most extended species of WT Ltn40 ($[D_{WT}+13H]^{13+}$). Unfortunately no PDB file exists

for W55D dimer, so direct comparisons cannot be made and instead comparisons are made with WT dimer (PDB 2JP1), which as discussed before omits 33 amino acids from the ID tail of each chain. Comparison of this theoretical CCS to the experimental CCS demonstrates that D_{W55D} populates a wide range of conformations more compact and extended than this theoretical value. The majority of observed species are more extended than the theoretical value, suggesting either the ID tail is extended out from the structural core and/or the whole structure is more flexible and therefore more capable of existing in a larger range of conformations where for some there will be a smaller interface region stabilising the dimer. The fact that this mutant occupies a wider range of charge states is also indicative of a more flexible system. Significantly, less dimer signal is observed for W55D in comparison to the WT Ltn and analysis of M_{W55D} suggests that these species are due to dissociated dimer and therefore D_{W55D} is considered less stable than D_{WT}. The likelihood of this would be increased if the dimer interface is indeed smaller and less stable in the W55D mutant. It is probable that the wide range of experimental CCS is due to both tail unfolding and increased flexibility in D_{W55D}.

In order to probe this hypothesis ECD studies were performed, as with WT dimer only the charge states of dimer where the net charge is odd were considered, namely $[D_{W55D}+9H]^{9+}$, $[D_{W55D}+11H]^{11+}$ and $[D_{W55D}+13H]^{13+}$. Analysis was performed considering zIc, c and z-type fragments as a product of dimer fragmentation (above axis histograms) or dissociation of dimer to monomer followed by fragmentation (below axis histograms). Both monomeric units are identical in sequence and therefore fragmentation maps are plotted onto a single chain (Figure 4.22 and Figure 4.23).

ECD of the lowest charge state ($[D_{w55D}+9H]^{9+}$) leads to extensive dissociation of dimer to monomer followed by extensive fragmentation along the backbone (Figure 4.22). A tendency for the lowest dimer charge state to dissociate to monomer was also observed for WT Ltn and attributed to either an encounter complex or a less stable, perhaps collapsed form of the dimer. Unlike WT Ltn, dissociation of dimer to monomer upon ECD fragmentation is observed for all charge states of W55D and again suggests that the dimer interface is not as strong in W55D, consistent with the studies of M_{W55D} which suggested that the dimer interface had been weakened and the high proportion of monomer observed is likely due to dissociation of dimer to monomer. These results highlight the significant structural influence point mutations can have, considering the position of the amino acid substitution in Ltn40 fold it is surprising that such a destabilisation is observed (Figure 4.24).

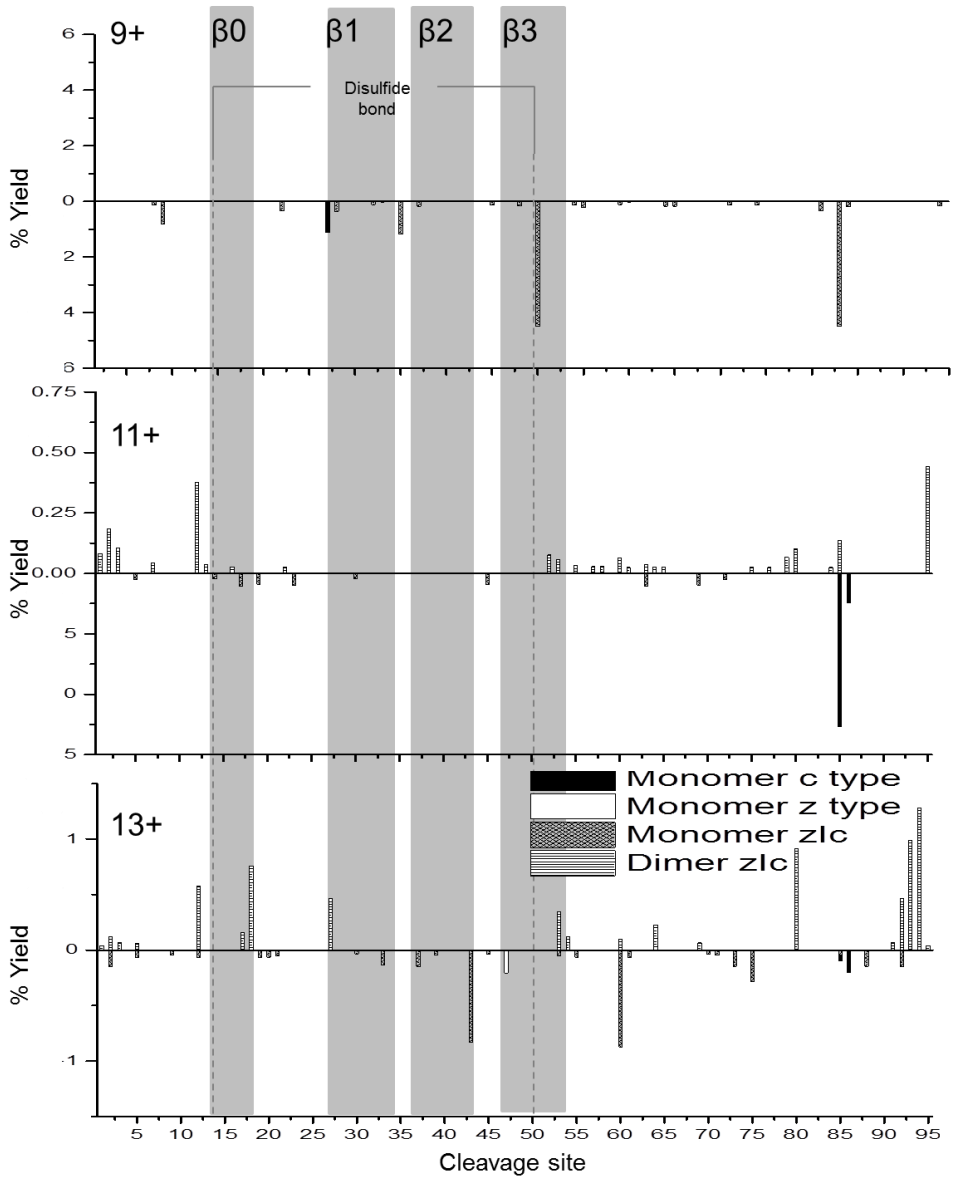

Figure 4.22: Percentage yields calculated for zIc, c- and z-type fragments as a function of cleavage site for dimeric W55D, charge states $[D+9H]^{9+}$, $[D+11H]^{11+}$ and $[D+13H]^{13+}$. Above axis histograms represent fragments formed while retaining the dimer interface. Below axis histograms represent fragments formed following dimer dissociation to monomer. For figure clarity the % yield scale is altered between charge states.

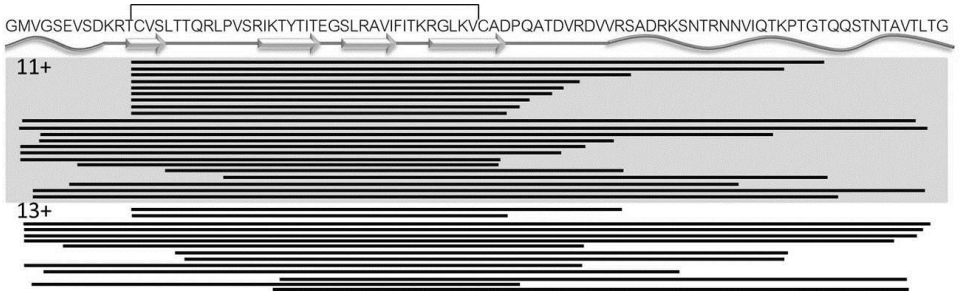

Figure 4.23: W55D sequence labelled with disulfide bond (solid line) and the secondary structural elements of Ltn40, annotated with all zIc fragments observed for dimeric W55D, where the dimer interface is retained.

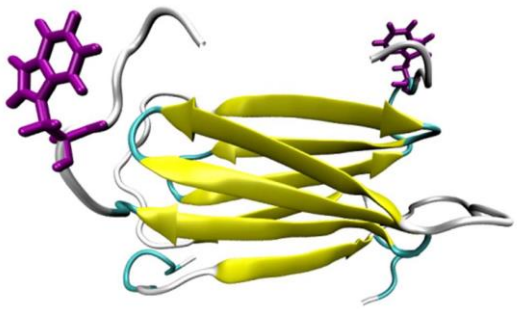

Figure 4.24: Structure of WT Ltn40 (PDB 1JP1) with residue W55 shown in purple, generated in VMD.

Interestingly, the highest intensity of fragmentation is observed for $[D_{W55D}+9H]^{9+}$, suggesting this species is the most structurally unstable conformation and can therefore undergo the greatest extent of fragmentation, consistent with the results of DT IM-MS experiments which suggest it is a collapsed form. For $[D_{W55D}+11H]^{11+}$ dimer fragmentation maps better with the Ltn40 fold, with the majority of fragmentation occurring at the termini and in flexible regions just before the β-sheet core (Figure 4.22 and Figure 4.23), suggesting that the dimeric species which survive the desolvation process are likely to be present in the Ltn40 fold. This finding is consistent with the studies of other Ltn constructs here, which highlighted that the solution folds of this metamorphic protein are retained in the gas-phase and can be probed using these techniques. For $[D_{W55D}+11H]^{11+}$ fragmentation is observed within the β0 strand, which increases in intensity for $[D_{W55D}+13H]^{13+}$. The increase in fragmentation along this region suggests it is the least stable of the β-strands and is the first to unfold from the structural core. It is therefore surmised here that the structural core of W55D in the Ltn40-like conformation unfolds from the N-terminus, interestingly the converse of Ltn10 which unfolds first from the C-terminus. ECD fragmentation maps allow the stability of each region to be ranked from least to most stable (β0<β1< β3< β2), which could help in the design of future mutants aimed to mimic this fold but with increased stability and decreased unfolding. For W55D clearly a lower proportion of dimer is present than for WT Ltn and therefore a more stable dimer mutant is required.

4.3.3.4 Visualising protein unfolding

By combining DT IM-MS results with ECD fragmentation maps for CC3 and W55D a detailed picture of the unfolding of both Ltn10 and Ltn40 can be built, giving insight into fold stability as a function of increased charge (Figure 4.25).

Using the methodology presented here, the secondary and tertiary structure of the mutants can be probed, enabling detailed examination of both of distinct folds of this metamorphic

protein. It is observed that CC3 is more structurally constrained than WT Ltn, as expected for a protein with an additional disulfide bond, and most likely retains in the Ltn10 fold in the gas-phase. The positioning of the additional disulfide bond pins the α-helix to the structural core and blocks the first unfolding pathway of Ltn10. For W55D the results suggest that the dimer interface is less stable in this mutant than in WT Ltn, leading to dissociation of dimer to monomer both upon ionisation and transfer and as a result of ECD fragmentation. The fragmentation maps of this mutant also suggest that for Ltn40 the first unfolding pathway involves unfolding of the β0 strand and N-terminus. The combination of these gas-phase techniques and their application to mutagenesis studies has potential to further the design of mutants with stabilised folds by highlighting susceptible regions.

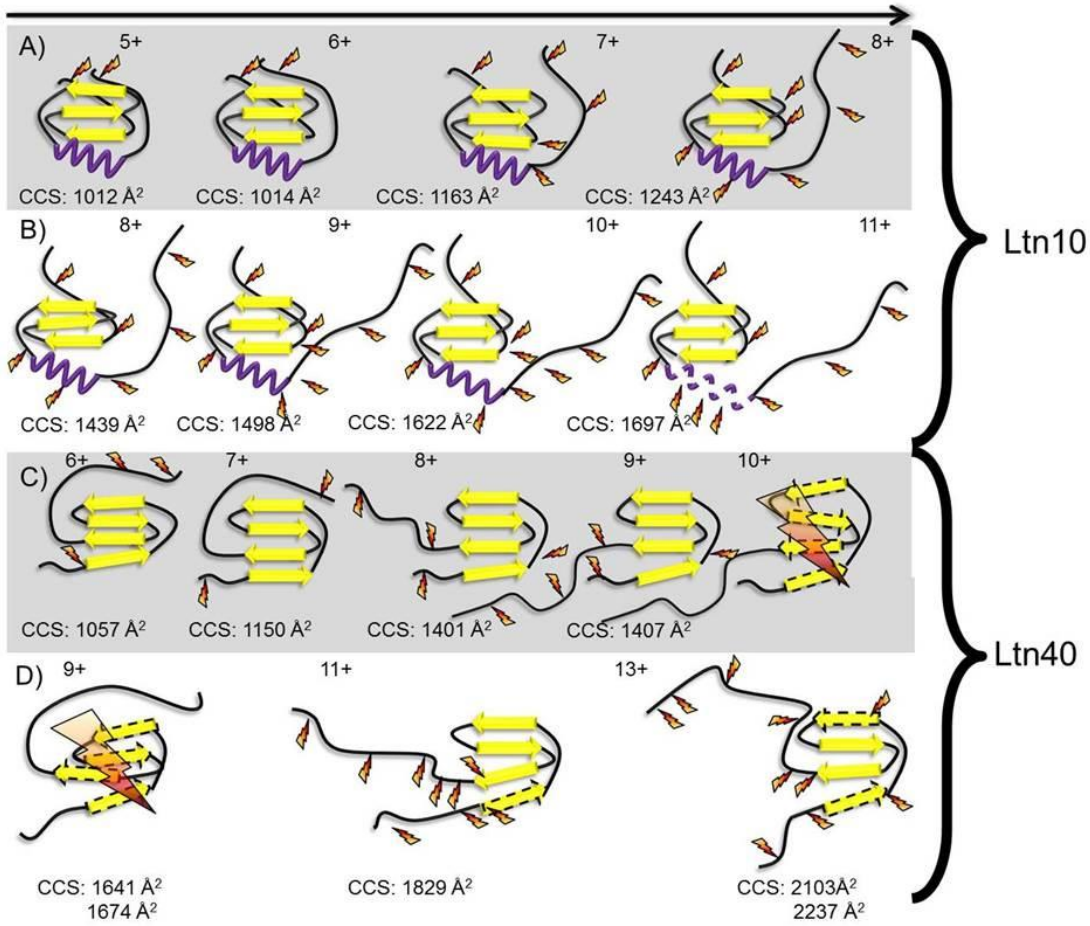

Figure 4.25: Cartoon representation of the unfolding of Ltn10 and Ltn40 folds as shown by DT IM-MS and ECD, probed through specific mutants, A) buffered CC3 B) supercharged CC3 C) monomeric W55D and D) dimeric W55D, where the dimer is represented here by a single monomeric unit. Yellow arrows represent β-sheet regions and purple spirals represent α-helical regions. Orange 'lightning bolts' indicate sites where ECD fragmentation is most significant, extensive fragmentation is depicted by a large lightning bolt and dashed lines represent a lack of defined structural elements.

4.4 Conclusions

The combination of DT IM-MS and ECD fragmentation can provide detailed insight into the fold and unfolding of proteins, with no constriction due to buffers or additives enabling multiple, dynamic conformations of a metamorphic protein to be studied simultaneously.

Results suggest that for WT Ltn in both its monomeric and dimeric forms the solution folds can be preserved in the gas-phase, over the time scales of the experiments present here (up to seconds), suggesting this combination of DT IM-MS and ECD has the potential to enable the study of multiple conformations of structurally dynamic proteins and has the opportunity to overcome many of the challenges associated with traditional structure elucidation tools. In both Ltn10 and Ltn40, the CCS in combination with ECD maps suggest that β-sheet regions may be preserved, with increased stability of these regions being observed. Further in-depth studies would be required to conclusively prove that these regions are indeed preserved, however, the results here suggest it is conceivable. Ltn would therefore be an excellent candidate for further gas-phase protein structural studies, to further study β-sheets in the gas-phase.

The analysis of the mutants CC3 and W55D provided further details on the distinct folds of Ltn. The results presented here confirm that CC3 is a suitable mimic for Ltn10, with the extra disulfide bond constricting and stabilising the fold. W55D, however, appears less stable as a dimer than WT Ltn and suggests the dimeric interface is weakened for this construct as compared with WT protein and signifying that future studies should consider the creation of a Ltn40 mutant with increased stability.

The ability of this combination of techniques to distinguish significant effects of point mutations on protein fold and unfolding highlights the wealth of information that can be obtained in this way. Information can be acquired on the relative stabilities of mutants in comparison to WT protein, and therefore has potential to be highly influential in the analysis of possible protein mimics of therapeutic potential. The power of using such techniques becomes evident when you consider the small sample volumes required for such studies, with as little as 25 μg of protein providing detailed insight into dynamic protein behaviour.

4.5 References

(1) Kuloglu, E. S.; McCaslin, D. R.; Kitabwalla, M.; Pauza, C. D.; Markley, J. L.; Volkman, B. F., *Biochemistry*, **2001**, *40*, 12486-12496.
(2) Handel, T.; Hamel, D. *Chemokines*; Elsevier: San Diego, 2009.
(3) Tuinstra, R. L.; Peterson, F. C.; Kutlesa, S.; Elgin, E. S.; Kron, M. A.; Volkman, B. F., *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105*, 5057-5062.
(4) Murzin, A. G., *Science*, **2008**, *320*, 1725-1726.
(5) Roche, S.; Rey, F. A.; Gaudin, Y.; Bressanelli, S., *Science*, **2007**, *315*, 843-848.
(6) Littler, D. R.; Harrop, S. J.; Fairlie, W. D.; Brown, L. J.; Pankhurst, G. J.; Pankhurst, S.; DeMaere, M. Z.; Campbell, T. J.; Bauskin, A. R.; Tonini, R., *J. Biol. Chem.*, **2004**, *279*, 9298-9305.
(7) Xu, M.; Arulandu, A.; Struck, D. K.; Swanson, S.; Sacchettini, J. C.; Young, R., *Science*, **2005**, *307*, 113-117.
(8) Mapelli, M.; Massimiliano, L.; Santaguida, S.; Musacchio, A., *Cell*, **2007**, *131*, 730-743.
(9) Sun, Q. Y.; Tyler, R. C.; Volkman, B. F.; Julian, R. R., *J. Am. Soc. Mass Spectrom.*, **2011**, *22*, 399-407.
(10) Breuker, K.; McLafferty, F. W., *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105*, 18145-18152.
(11) Jurneczko, E.; Barran, P. E., *Analyst*, **2011**.
(12) Myung, S.; Badman, E. R.; Lee, Y. J.; Clemmer, D. E., *J. Phys. Chem. A*, **2002**, *106*, 9976-9982.
(13) Konermann, L.; Pan, J.; Liu, Y.-H., *Chem. Soc. Rev.*, **2011**, *40*, 1224-1234.
(14) Vahidi, S.; Stocks, B. B.; Liaghati-Mobarhan, Y.; Konermann, L., *Anal. Chem.*, **2013**, *85*, 8618-8625.
(15) Pan, Y.; Brown, L.; Konermann, L., *J. Am. Chem. Soc.*, **2011**, *133*, 20237-20244.
(16) Hambly, D. M.; Gross, M. L., *J. Am. Soc. Mass Spectrom.*, **2005**, *16*, 2057-2063.
(17) Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L., *Anal. Chem.*, **2009**, *81*, 6563-6571.
(18) Chen, J.; Rempel, D. L.; Gross, M. L., *J. Am. Chem. Soc.*, **2010**, *132*, 15502-15504.
(19) Chen, J.; Cui, W.; Giblin, D.; Gross, M. L., *J. Am. Soc. Mass Spectrom.*, **2012**, *23*, 1306-1318.
(20) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., *J. Am. Chem. Soc.*, **1995**, *117*, 10141-10142.
(21) Badman, E.; Hoaglund-Hyzer, C. S.; Clemmer, D. E., *Anal. Chem.*, **2001**, *73*, 6000-6007.
(22) Smith, D. P.; Giles, K.; Bateman, R. H.; Radford, S. E.; Ashcroft, A. E., *J. Am. Soc. Mass Spectrom.*, **2007**, *18*, 2180-2190.
(23) Wyttenbach, T.; Grabenauer, M.; Thalassinos, K.; Scrivens, J. H.; Bowers, M. T., *J. Phys. Chem. B*, **2009**, *114*, 437-447.
(24) Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T., *Nat. Chem.*, **2011**, *3*, 172-177.
(25) Jenner, M.; Ellis, J.; Huang, W. C.; Lloyd Raven, E.; Roberts, G. C.; Oldham, N. J., *Angew. Chem.*, **2011**, *50*, 8291-8294.
(26) Skinner, O.; McLafferty, F.; Breuker, K., *J. Am. Soc. Mass Spectrom.*, **2012**, *23*, 1011-1014.
(27) Wyttenbach, T.; Bowers, M. T., *J. Phys. Chem. B*, **2011**, *115*, 12266-12275.
(28) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., *J. Am. Chem. Soc.*, **1998**, *120*, 3265-3266.
(29) Breuker, K.; Oh, H. B.; Lin, C.; Carpenter, B. K.; McLafferty, F. W., *Proc. Natl. Acad. Sci. U. S. A.*, **2004**, *101*, 14011-14016.
(30) Breuker, K.; Brüschweiler, S.; Tollinger, M., *Angew. Chem.*, **2011**, *50*, 873-877.
(31) Breuker, K.; Oh, H. B.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W., *J. Am. Chem. Soc.*, **2002**, *124*, 6407-6420.
(32) Horn, D. M.; Breuker, K.; Frank, A. J.; McLafferty, F. W., *J. Am. Chem. Soc.*, **2001**, *123*, 9792-9799.
(33) Cooper, H. J.; Håkansson, K.; Marshall, A. G., *Mass Spectrom. Rev.*, **2005**, *24*, 201-222.
(34) Oh, H.; Breuker, K.; Sze, S. K.; Ge, Y.; Carpenter, B. K.; McLafferty, F. W., *Proc. Natl. Acad. Sci. U. S. A.*, **2002**, *99*, 15863-15868.
(35) Moss, C. L.; Chamot-Rooke, J.; Nicol, E.; Brown, J.; Campuzano, I.; Richardson, K.; Williams, J. P.; Bush, M. F.; Bythell, B.; Paizs, B.; Turecek, F., *J. Phys. Chem. B*, **2012**, *116*, 3445-3456.
(36) Kalapothakis, J. M. D.; Berezovskaya, Y.; Zampronio, C. G.; Faull, P. A.; Barran, P. E.; Cooper, H. J., *Chem. Comm.*, **2014**.

(37) Peterson, F. C.; Elgin, E. S.; Nelson, T. J.; Zhang, F.; Hoeger, T. J.; Linhardt, R. J.; Volkman, B. F., *J. Biol. Chem.*, **2004**, *279*, 12598-12604.
(38) Tuinstra, R. L.; Peterson, F. C.; Elgin, E. S.; Pelzek, A. J.; Volkman, B. F., *Biochemistry*, **2007**, *46*, 2564-2573.
(39) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.; Luo, R. C. W., W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg,; G. Seabra, I. K., K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell,; T. Steinbrecher, H. G., Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H.; Mathews, M. G. S., C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and; Kollman, P. A. *AMBER 11*; University of California: San Francisco, 2010
(40) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., *Proteins: Struct. Funct. Bioinf.*, **2010**, *78*, 1950-1958.
(41) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., *J. Phys. Chem. A*, **1996**, *100*, 16082-16086.
(42) Heck, A. J. R., *Nat Meth*, **2008**, *5*, 927-933.
(43) Chowdhury, S. K.; Katta, V.; Chait, B. T., *J. Am. Chem. Soc.*, **1990**, *112*, 9012-9013.
(44) Roepstorff, P.; Fohlman, J., *Biomed. Mass Spectrom.*, **1984**, *11*, 601.
(45) Iavarone, A. T.; Paech, K.; Williams, E. R., *Anal. Chem.*, **2004**, *76*, 2231-2238.
(46) Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W., *J. Am. Chem. Soc.*, **1999**, *121*, 2857-2862.
(47) Edmond de Hoffmann; Stroobant, V. *Mass Spectrometry: Principles and Applications*; 3 ed.; John Wiley and Sons: Chichester, 2007.
(48) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., *J. Am. Chem. Soc.*, **1997**, *119*, 2240-2248.
(49) Iavarone, A. T.; Williams, E. R., *Int. J. Mass. Spectrom.*, **2002**, *219*, 63-72.
(50) Sterling, H.; Kintzer, A.; Feld, G.; Cassou, C.; Krantz, B.; Williams, E., *J. Am. Soc. Mass Spectrom.*, **2012**, *23*, 191-200.

5

Investigating interactions with GAGs: Lymphotactin

Glycosaminoglycan (GAG) binding is essential for chemokine activity. Chemokine:GAG interactions mediate binding of the chemokine to the vasculature and enable a chemokine concentration gradient to be established, a process important for cell migration. Understanding more about this intricate biological process, the complexes formed and their conformations is vital for a better understanding of chemotaxis. This chapter focuses on the chemokine lymphotactin and its ability to bind a model GAG (fondparinux). A number of specific mutants, designed either to stabilise one of the distinct folds of lymphotactin or to destabilise lymphotactin:GAG binding, are also considered. In this chapter MS, DT IM-MS and ECD are utilised to identify the lymphotactin:GAG complexes formed. Extensive aggregation was observed upon addition of fondparinux to lymphotactin solutions and TEM enables these larger aggregates to be visually examined.

5.1 Introduction

5.1.1 Chemokine:GAG binding

Proteins involved in immune modulation, such as cytokines and growth factors, fulfil their biological function through high-affinity interactions with cell-surface receptors[1]. Recent studies, however, have shown that many of these proteins also exhibit low-affinity interactions with glycosaminoglycans (GAGs) or proteoglycans, proteins to which GAG chains are attached[2]. GAGs are long unbranched polysaccharide chains which consist of repeating disaccharide units and are extremely variable in nature, with sequences which differ in the basic composition of the saccharide, linkage, chain length, acetylation and N- and O-sulfonation[3]. They can, however, be divided into six major families; heparin, heparin sulphate, chondroitin sulphate, dermatan sulphate, keratin sulphate and hyaluronic acid[4], based on the disaccharide linkage. Heparin sulfate is the most ubiquitous, being expressed on virtually every cell in the body and comprising 50 % to 90 % of endothelial proteoglycans[5].

In recent years it has become apparent that in order to fulfil their biological function chemokines are required to interact both with their G protein coupled receptor (GPCR) and to GAGs on cell surfaces or in the extracellular matrix[2,6]. The ability of GAGs to bind chemokines is anticipated from a simple electrostatic point of view, with GAGs being highly acidic and chemokines generally being basic[7]. There exist, however, two acidic chemokines of the CC subfamily, both of which have been shown to bind GAGs, signifying that chemokine:GAG interactions display specificity and are not solely dominated by non-specific electrostatic interactions[8]. Chemokines have been shown to bind GAGs as a localisation mechanism, in order to act as a directional signal for migrating cells. The absence of such interactions could cause rapid diffusion of chemokines to levels below the threshold for receptor activation, particularly in the presence of vascular flow. Moreover, it has been suggested that interactions with GAGs confer specificity to this complex regulatory system and it has been proposed that chemokines could be directed to different locations based on varying affinities for different GAGs; an important point to consider in combination with the fact that many chemokines bind the same receptor with similar high affinity[9-11].

The functional importance of such interactions was solidified through an influential study into the CC class of chemokines by Proudfoot *et al*[8], in which site-specific mutations were implemented in order to inhibit GAG binding and were shown *in vitro* to minimally perturb receptor binding. *In vivo,* however, these mutants were ineffective at inducing cell migration

in comparison to the WT chemokine, attributed to the necessity of chemokine:GAG interactions. Studies into viral mechanisms of immunomodulation have also highlighted the importance of chemokine:GAG interactions through the discovery that a number of viruses produce chemokine binding proteins, some of which function by blocking the GAG binding epitopes[12-15].

Chemokine:GAG binding is an intricate process and understanding more about these pathways, complexes formed and their conformations is essential. Such studies are challenging due to the extreme heterogeneity of GAGs and difficulties in producing them recombinantly, hence most studies initially focus on model GAGs, generally of the heparin/heparin sulphate subfamilies due to their commercial availability. A number of approaches have been presented to-date to gain insight into these important interactions including: heparin affinity chromatography, titrated heparin binding assays, surface plasmon resonance, *in vivo* cellular recruitment and NMR studies[16]. A small number of particularly sophisticated and detailed studies have been presented to solve the structures of complexes formed between chemokines and GAGs and to identify key residues involved in binding[17,18]. Furthermore, the common sequence motif BBXB has been identified as responsible for GAG binding, where B is a basic amino acid[19].

Chemokines generally exist in solution as monomers or dimers and occasionally tetramers[2]; however, upon GAG binding they have been found, and predicted, to exist in primarily dimeric or tetrameric GAG-bound forms, although in some cases monomer:GAG complexes have been observed[20,21]. Detailed structural studies of such complexes have allowed two chemokine:GAG crystal structures to be solved; CCL5 and CXCL2, both in complexes with a heparin-derived disaccharide and both of which form GAG bound dimers[22,23]. In these studies attempts were also made to crystalize these proteins with longer oligosaccharides; however, they were unsuccessful due to problems associated with aggregation and a corresponding inability to produce high quality crystals. Heparin has been shown *in vitro* to induce higher-order aggregates in solution[11]. This ability to produce higher order oligomers in the presence of GAGs is deemed essential for the activity of particular chemokines *in vivo*[8] and thought to help over-come issues associated with vascular flow[24].

Recently mass spectrometry (MS) has emerged as an alternative and valuable method to study these interactions; MS does not require labels and allows low intensity bound species to be separated and identified. Furthermore MS is well suited to the study of aggregating systems[25-28]. MS can be used to determine the stoichiometry of chemokine:GAG interactions or in combination with proteolytic digests to identify binding sites[29]. The development and

application of MS to study chemokine:GAG interactions has been pioneered through research from Julie Learys' group, who have produced a number of sophisticated studies in this field[30-33]. One such study used MS to determine the stoichiometry of GAG-bound complexes of a number of chemokines from the CC subfamily (CCL2, CCL8, CCL7, CCL11, CCL13 and CCL27), finding that members of this subclass could be differentiated by their GAG binding, with complexes presenting as either monomers or dimers[34]. More recently the application of ion mobility-mass spectrometry (IM-MS) in combination with collision induced dissociation (CID) studies has been presented to study the conformations and corresponding stabilities of chemokine:GAG complexes. Results of such studies demonstrated a stabilising effect upon GAG binding to CCL2, the extent of which can differ with the sulfonation pattern of the GAG[35,36]. Furthermore, these investigations have shown that complexes formed and their relative stabilities can differ depending on the class of chemokine studied, with the CC and CXC families of chemokine exhibiting different behaviour[36]. This work highlights the detail such studies can provide, particularly in combination with other techniques.

5.1.2 Ltn:GAG binding

The metamorphic chemokine lymphotactin (Ltn) is thought to segregate the two functional roles of chemokines through a conformational barrier. The conserved chemokine fold (Ltn10) has been found to be responsible for binding the relevant GPCR, namely XCR1[37], whilst the dimeric fold (Ltn40) is thought to be inactive against XCR1[38]. The GAG binding functionality of Ltn has been assessed *via* elution from a heparin-sepharose column with a sodium chloride gradient[38,39]. WT Ltn was found to elute in two broad fractions in ~450 and 700 mM NaCl which were assigned as the two conformations (Ltn10 and Ltn40 respectively) of WT Ltn, having significantly different binding affinities. This hypothesis was confirmed through the study of Ltn10 and Ltn40 mutants (CC3 and W55D respectively)[1,38]. These findings suggest that although both folds can bind heparin, Ltn40 encodes for high affinity GAG binding. Furthermore, the addition of heparin to a solution of Ltn can shift the equilibrium towards Ltn40[39].

Despite the higher GAG binding affinity of the Ltn40 conformation, to-date detailed studies have focused on Ltn10 binding. This is principally due to the fact that such studies were carried out prior to the dimeric structure of Ltn being solved[38] and hence were performed under conditions in which the equilibrium is strongly shifted towards the monomeric form[40,41]. Considering the propensity of GAGs to bind to basic amino acids, Peterson *et al*[41] conducted a study in which most of the lysine and arginine residues were substituted with

alanine to identify key amino acids for GAG binding. The studies showed that the greatest decrease in heparin binding affinities occurred for substitutions at R23A and R43A, with a maximum decrease in affinity determined for the double point mutant. It is important to note that in WT Ltn, charge repulsion between R23 and R43 is thought to destabilise the Ltn10 fold, an effect which is overcome through addition of salt. Substitution for alanine at either R23 or R43 reduces charge repulsion, stabilising Ltn10, and pushing the equilibrium towards the monomeric fold. The Ltn10 and Ltn40 three-dimensional folds are, however, unperturbed by these substitutions.

This chapter focuses on the use of MS in combination with DT IM-MS and transmission electron microscopy (TEM) to study the interactions of a model GAG, fondaparinux (Fx), with WT Ltn and mutants designed to stabilise Ltn10 (CC3) and Ltn40 (W55D). Fondaparinux (also known as Atrixia) is a pentameric GAG (Figure 5.1) chemically related to heparin, which has been marketed by GlaxoSmithKline as an anticoagulant and has been used as a treatment for deep vein thrombosis[42]. It serves as an excellent model GAG due to its availability in an extremely pure form with a well characterised and controlled sulfonation pattern and hence has been used in previous studies of chemokine:GAG binding[43]. Previous NMR-based GAG binding studies of Ltn highlighted that pentameric GAGs can bind Ltn10 with low affinity (K_d ~100 μM), however, upon addition of longer chain length GAGs Ltn solutions exhibit greater aggregation and form insoluble complexes[41] making Fx an attractive model for initial studies. Additional MS and TEM studies were performed to study interactions of Fx with single (R23A and R43A) and double (R23/R43A) point mutants designed to have lower affinity towards GAGs. The studies presented here aim to provide insight into the interactions of Ltn with a model GAG, in order to definitively determine the stoichiometry of interactions and to gain understanding into the conformations of these complexes. Such information would provide insight into this complicated, biologically important pathway.

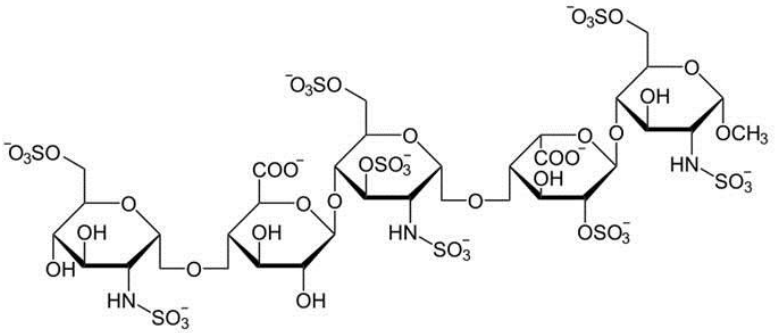

Figure 5.1: Structure of fondaparinux

5.2 Experimental

5.2.1 Protein and glycosaminoglycan samples

All recombinant human lympohactin (Ltn) samples were expressed and purified by the Volkman Research Group (Medical College of Wisconsin, Milwaukee, USA) as previously described[37,40,41]. It is important to note due to differences in the expression and purification systems used for WT Ltn and the mutants CC3 and W55D, residues M63 and M73 were mutated to valine and alanine respectively in both CC3 and W55D. Proteins were received as lyopholised samples and stored at -20 °C for subsequent study. Stock solutions of lymphotactin samples were typically prepared at 100, 150 or 200 μM in 20 mM ammonium acetate (AmAc) and stored in aliquots at -20 °C until use. Concentrations were checked using the Pierce BCA concentration assay (Thermo Fischer Scientific, Wilmington, DE, USA) following the manufacturer's instructions but using volumes 10 times lower than stated in the manual, afforded through the use of low volume UV cuvettes.

The pentameric glycosaminoglycan (GAG) foundaparinux sodium was a gift from GlaxoSmithKline (Brentford, UK). Foundaparinux was prepared at a stock concentration of 587 μM in water. Aliquots of which were dialysed overnight versus water using a 1,000 molecular weight cut-off Micro DispoDialyzer (Harvard Apparatus, MA, USA), and stored at -20 °C until use.

5.2.2 Sample preparation for MS studies

In all cases samples were prepared in 20 mM AmAc, at pH 6.8 and analysed immediately after sample preparation.

For all Ltn samples individual protein plus Fx solutions were prepared at 50 μM protein concentration at a 1:1 ratio of Ltn:Fx. WT Ltn was additionally studied at a 1:1 ratio at 25 and 10 μM, and at a 2:1 ratio (Ltn:Fx) at a 25 μM Ltn concentration. WT 1-72 was also studied at 1:1 and 2:1 ratios (Ltn:Fx) at a 25 μM protein concentration.

The interactions of WT 72-93 plus WT 1-72 with Fx were also considered using MS. Here sample solutions were prepared WT 1-72: WT 72-93: Fx ratios of 1:1:1 (50 μM), 2:2:1 (50:50:25 μM), 2:2:1 (50:50:25 μM) and 4:4:1 (25:25:6.25 μM).

All MS studies were performed on a Q-TOF Ultima (Waters, Manchester, UK), for details on the instrument and its operating conditions see Chapter 2 section 2.2. Collision induced

dissociation studies, experiments involving Ltn 72-93 and all experiments on R23A, R43A and R23A/R43A were performed post high mass upgrade.

5.2.3 Sample preparation for DT IM-MS studies

For determination of CCS for Ltn:Fx species WT, CC3 (monomer mutant) and W55D (dimer mutant) constructs were prepared at 100 μM protein concentration in 20 mM AmAc at a 1:1 (Ltn:Fx) ratio. WT 1-72 was studied at 50 μM in 20 mM AmAc at a 2:1 ratio with fondparinux.

All DT IM-MS measurements were performed on an in-house modified Q-TOF (Micromass UK Ltd.), described in Chapter 2 section 2.3. For all DT IM-MS studies the source was held at 100 °C to aid desolvation.

5.2.4 Sample preparation for ECD studies

For WT:Fx binding interface studies by ECD, samples were prepared at a 1:1 ratio at 30 μM in 100 mM AmAc, pH 6.8. All ECD studies were performed on a 12T Apex Ultra Qh FT-ICR mass spectrometer (Bruker Daltonik Gmbh, Bremen, Germany), details on the instrument and its operation can be found in Chapter 2 section 2.5.

5.2.5 Sample preparation for TEM studies

For transmission electron microscopy (TEM) studies WT, CC3, W55D, WT 1-72, WT 72-93, R23A, R43A and R23A/R43A samples were prepared at 50 μM protein concentration at a 1:1 ratio of Ltn:Fx. WT Ltn was additionally studied at 10 μM protein concentration at a 1:1 ratio of Ltn:Fx. Furthermore, a sample containing WT 1-72 plus WT 72-93 plus Fx present at a 1:1:1 ratio (50:50:50 μM) was also studied. In all cases samples were prepared in 20 mM AmAc at pH 6.8 and left to incubate for 30 minutes before grid preparation. In addition individual solutions of WT Ltn and Fx, at 50 μM in 20 mM AmAc were studied as controls. For details on TEM grid preparation see Chapter 2 section 2.7.

All TEM experiments were performed on a Philips CM120 Biotin transmission electron microscope (Philips, Eindhoven, Netherlands), details on the instrument and its operation can be found in Chapter 2 section 2.7.

5.3 Results and discussion

5.3.1 Interactions with WT Ltn

Firstly the interactions between Fx and WT Ltn were considered. This was of interest due to the metamorphic nature of WT Ltn, which exists in equilibrium between the conformational and functional forms, each of which is thought to have different affinity for GAGs.

5.3.1.1 Stoichiometry of binding: insights from MS

The spectrum obtained from a 1:1 mixture at 50 μM (Figure 5.2B) contains a number of peaks due to Fx bound to Ltn over a range of charge states. Fx binds to both the monomeric form of Ltn (M_{WT}+Fx) and the dimeric form; dimeric WT Ltn is capable of binding either one (D_{WT}+Fx) or two (D_{WT}+2Fx) Fx chains, with expected and observed masses of all species being in good agreement (Appendix 4 Table A4.1). Previous heparin-sepharose column binding studies of WT Ltn showed that Ltn eluted in two fractions with significantly different affinities for heparin, assigned to monomer and dimer interactions, with the dimer fraction being larger. It was therefore hypothesised that here WT Ltn would be observed to bind Fx as a monomer to a lesser extent than the dimer. Surprisingly, the intensity of M_{WT}+Fx is higher than anticipated, with $[M_{WT}+Fx+6H]^{6+}$ the most intense Fx-bound species. It is possible that this species is formed faster or is an encounter species *en route* to a dimer-bound species. It is not possible, however, to test this theory with MS, as upon addition of Fx to WT Ltn extensive aggregation occurs, with large aggregates visible in the n-ESI tip (example images are shown in Appendix 4 Figure A4.1). The solution changes from clear to cloudy in appearance and is accompanied by a significant reduction in signal intensity. The higher order aggregation induced upon addition of Fx is studied further using transmission electron microscopy (TEM) in section 5.3.1.4. This extremely rapid and extensive aggregation means it is not possible to perform any MS time course experiments to determine how bound and unbound species vary as a function of time.

High concentrations of binding partners can lead to non-specific interactions, therefore the specificity of Ltn to form monomeric and dimeric Fx-bound species is assessed as a function of concentration. In order to achieve this, sample solutions were prepared at a protein concentration of 25 and 10 μM at 1:1 ratio with Fx (Figure 5.2C and D). In all cases monomeric and dimeric Ltn are observed to bind to Fx and even more significantly the monomeric species is the most intense in all cases. The spectra obtained for the 1:1 mixtures at 25 and 10 μM have an obvious raised baseline, which is attributed to extensive

aggregation occurring for all samples. The lower concentration of unaggregated material in these samples leads to spectra of low signal-to-noise and hence raised baselines.

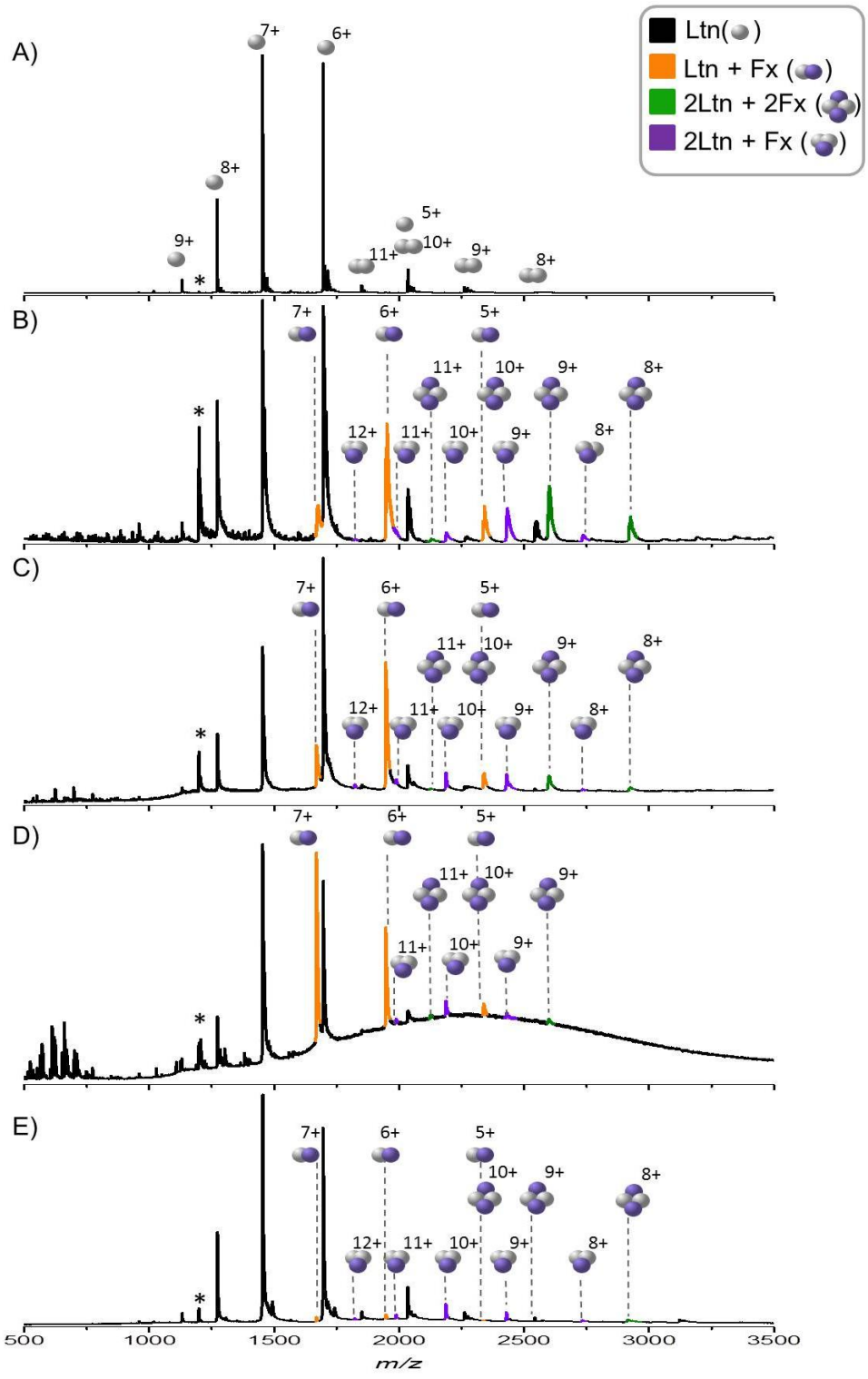

Figure 5.2: Spectra obtained for WT Ltn A) 50 µM B) 50 µM plus 50 µM Fx (1:1, Ltn:Fx) C) 25 µM plus 25 µM Fx (1:1) D) 10 µM plus 10 µM Fx (1:1) and E) 25 µM plus 12.5 µM Fx (2:1) * denotes a species observed in all spectra presumably a contaminant left over after purification. In panel A all Ltn species are labelled. For figure clarity in panels B-E only GAG bound species are labelled.

The presence of M_{WT}+Fx species at even the lowest concentration suggests these complexes are not just a product of non-specific, high concentration interactions and instead there is specificity to their formation. To explore this further, a sample was prepared at a 2:1 at 25 μM Ltn concentration (Figure 5.2E). This shows a lower but still significant proportion of M_{WT}+Fx species, highlighting that although formation of monomeric Fx bound complexes can be limited by increasing Ltn concentration relative to Fx concentration, under all conditions and concentrations employed here the monomeric Fx bound species persists. Furthermore, in all cases the 7+ and 6+ charge states are the most intense unbound Ltn species, consistent with the findings in the absence of Fx (Figure 5.2A) and demonstrating Fx is not significantly perturbing the equilibrium or charge state distribution. Additional peaks at low *m/z* are observed in all cases due to contaminants and salts present in the samples.

The stability of both $[M_{WT}+Fx+5H]^{5+}$ and $[M_{WT}+Fx+6H]^{6+}$ were probed through collision induced dissociation (CID) experiments, with the first and highest intensity fragmentation pathway being loss of SO_3 from Fx (a common fragmentation pathway of Fx in isolation[44]), with the M_{WT}+Fx interaction remaining intact (Appendix 4 Figure A4.2). The fact that Fx fragments before the non-covalent complex dissociates reinforces the fact that strong, stable interactions are occurring here and suggests that M_{WT}+Fx species are indeed stable, specific complexes.

The propensity of Fx to bind specifically to monomeric Ltn in addition to dimeric Ltn suggests that despite its structural plasticity, when considering GAG binding the functions of the two distinct folds are not as separate and defined as previously thought. Dimeric Ltn was previously considered to be responsible for high affinity binding; however, here, monomeric Ltn clearly binds with high intensity. Other chemokines are known to bind GAGs as both monomers and dimers[43], suggesting that Ltn could behave in a similar fashion to the 'traditional' structurally constrained chemokines, capable of binding GAGs in a number of oligomeric states.

5.3.1.2 Conformations of WT:Fx complexes: DT IM-MS studies

Due to the weak signal obtained for the Ltn plus Fx samples, due to aggregation of a significant proportion of the material into large unresolvable complexes, the source pressure and voltages on the DT IM-MS instrument were tuned to preserve Fx-bound species. This in turn led to a reduction in signal for unbound protein and an inability to observe some unbound charge states. Therefore, for comparisons between unbound and bound species,

CCS for the unbound species (Figure 5.3) were determined in the absence of Fx, and are as reported in Chapter 4.

To determine if the presence of Fx can affect the conformations of unbound protein, the charge states observed in the presence of Fx were analysed in all cases. CCS determined for all unbound WT Ltn species determined in the absence and presence of Fx are shown in Table 5.1, and were found to remain the same within experimental error, suggesting that Fx does not affect the conformation of unbound WT Ltn in solution (in contrast to the findings of c-MYC:MAX upon addition of the ligand 10058-F4 in Chapter 3).

Species	***CCS/ Å²***	
	Absence of Fx	***Presence of Fx***
$[M_{WT}+6H]^{6+}$	1021 ± 27	1034 ± 47
$[M_{WT}+7H]^{7+}$	1100 ± 63	1077 ± 34
$[D_{WT}+12H]^{12+}$	1894*	1893 ± 21

Table 5.1: Experimentally determined average CCS for all unbound WT species observed in the absence and presence of Fx, the error is reported as the standard deviation between repeats. *species observed in all repeats but only resolvable in one.

Considering now the Fx-bound species (Figure 5.3 and Appendix 4 Table A4.2), in MS experiments $[M_{WT}+Fx+5H]^{5+}$ and $[M_{WT}+Fx+6H]^{6+}$ were the highest intensity bound species, suggesting that formation of these species is favourable. DT IM-MS experiments show that despite the additional mass from Fx, $[M_{WT}+Fx+5H]^{5+}$ adopts a conformation which has a smaller CCS than the unbound species at the same charge state and $[M_{WT}+Fx+6H]^{6+}$ is within error of the unbound species. This suggests that there has been a conformational tightening upon Fx binding, supporting the idea that these species are due to stable, specific binding.

The CCS of $[M_{WT}+Fx+7H]^{7+}$, however, is significantly larger than the CCS of both its corresponding unbound counterpart ($[M_{WT}+7H]^{7+}$) and the bound species of lower charge states, with a ~50 % increase in CCS for $[M_{WT}+Fx+7H]^{7+}$ compared with $[M_{WT}+Fx+5H]^{5+}$ (1247 vs 845 $Å^2$). This large increase in CCS across the charge states suggests either that the 7+ bound complex is non-specific and/or that binding of Fx can produce flexible species capable of adopting a wide range of conformations. Considering the inherent flexibility of Ltn, it is likely that this protein could form complexes with different structures.

Binding of either one or two units of Fx to a dimeric form of Ltn produces compact complexes with CCS within error or smaller than those of the unbound dimer. Due to the additional mass from Fx, larger CCS could be expected, however, CCS here suggest that there is a conformational tightening upon binding of Fx for dimeric Ltn. For both D_{WT}+Fx

and D_{WT}+2Fx, CCS of the different charge states remain fairly constant increasing by only 5 % and 16 %, respectively. In both cases therefore a stable conformational family is formed upon binding of Fx, which is not significantly affected by charge. Furthermore, CCS obtained for D_{WT}+Fx and D_{WT}+2Fx are very similar, with the dimer bound to two Fx chains being only marginally larger than the dimer bound to a single Fx at the same charge state.

Taken together, these results suggest that WT Ltn can bind Fx specifically, producing stable compact conformations in both monomeric and dimeric forms, with the dimer capable of binding one or two Fx chains specifically and with monomeric conformations appearing more flexible than dimeric conformations.

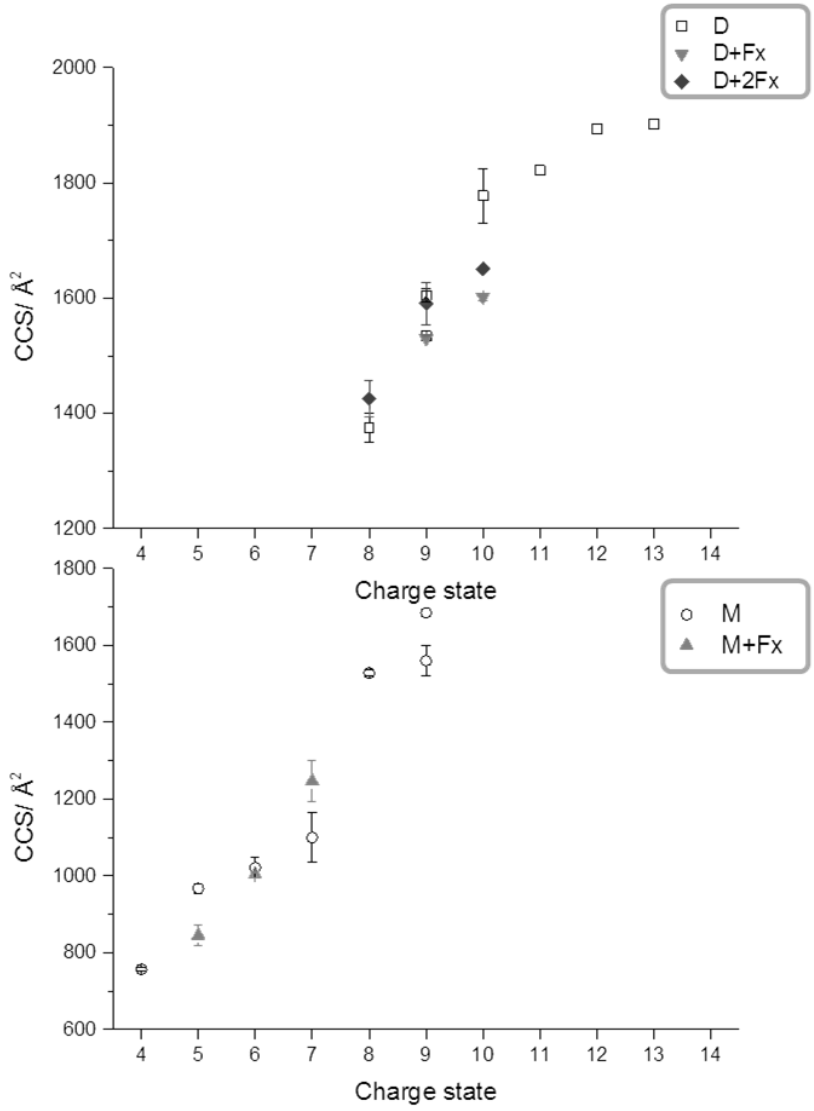

Figure 5.3: Average CCS of WT Ltn and Fx bound Ltn, dimeric (top) and monomeric (bottom) species, calculated from three repeats, error bars represent standard deviation between these values. Unbound species were determined in the absence of Fx (hollow symbols). Filled symbols represent Fx bound species.

5.3.1.3 Investigating the WT:Fx binding interface through ECD fragmentation.

In order to further examine the binding of Fx to WT Ltn, both the M_{WT} +Fx and the D_{WT}+Fx species were subjected to ECD fragmentation.

Considering M_{WT}+Fx first, charge states 5-7+ were examined, all of which exhibit very little fragmentation, with the 5+ charge state showing no visible fragment ions, further emphasising the stability of these monomeric bound species. $[M_{WT}+Fx+6H]^{6+}$ remains mostly intact and fragments only at the last two amino acids of the C-terminus following ECD. Both the 5+ and 6+ unbound species could not be isolated at high enough intensity for ECD fragmentation and hence comparisons with unbound species cannot be made here. $[M_{WT}+Fx+7H]^{7+}$ undergoes greater fragmentation, consistent with this species having a larger CCS and suggesting that it is more open and would have more potential sites of cleavage available. Indeed after being subjected to ECD, dissociation of $[M_{WT}+Fx+7H]^{7+}$ to form M_{WT} and fragments corresponding to M_{WT} with cleavage at the N-terminus are observed, suggesting Fx binding at this charge state is not as strong. A number of fragments are observed corresponding to retention of the M_{WT}+Fx complex, with the most intense being due to fragmentation of the first amino acid at the N-terminus or final amino acid at the C-terminus. When subjected to ECD $[M_{WT}+7H]^{7+}$ showed limited fragmentation, resulting in loss of the α-helix and ID tail (Chapter 4 section 4.3.2) which is not observed here for $[M_{WT}+Fx+7H]^{7+}$, suggesting these regions are more stabilised or less available for cleavage. Additionally, there are a number of low intensity, low molecular weight fragments which could not be attributed to c- or z-type fragments of monomer but could, however, be due to internal fragments of WT Ltn or Fx bound to internal fragments of WT Ltn. No single species of unique molecular weight that could be conclusively identified as either were observed.

The limited fragmentation along the sequence of WT Ltn coupled with no observable loss of the ID tail suggests that the tail is most likely involved in stabilising the complex, consistent with findings from DT IM-MS which show compact species. Additionally all species show a number of peaks due to electron capture with no dissociation resulting in charge reduction of M_{WT}+Fx and lower intensity peaks due to charge reduction of the *m/z* coincident D_{WT}+2Fx species. Accurate mass measurements and the presence of these species show that for all peaks assigned as M_{WT}+Fx and studied here a small proportion of D_{WT}+2Fx exists as well, however, M_{WT}+Fx is the dominant species in each case.

Due to their high stability and limited fragmentation, the precise binding site cannot be mapped for the M_{WT}+Fx species using ECD fragmentation. This could, however, also be an indication that Fx binds to a basic patch across the surface of monomeric Ltn, creating an extended binding interface, stabilised through multiple interactions. Figure 5.4A highlights all basic amino acids in Ltn10, the surface of monomeric Ltn contains large basic patches which, if Fx does bind here, would explain the lack of fragmentation along the backbone. The unbound species in ECD experiments $[M_{WT}+7H]^{7+}$ fragmented, resulting in a complete loss of the α-helix and ID tail (Chapter 4 section 4.3.2), however, the main basic patch on Ltn10 encompasses both the β-sheet core and α-helix, and if Fx binding does occur over this region would prevent this fragmentation pathway, consistent with experimental observations. In contrast to the unbound WT Ltn, the ID tail is not lost in M_{WT}+Fx suggesting it may also be involved in binding or complex formation.

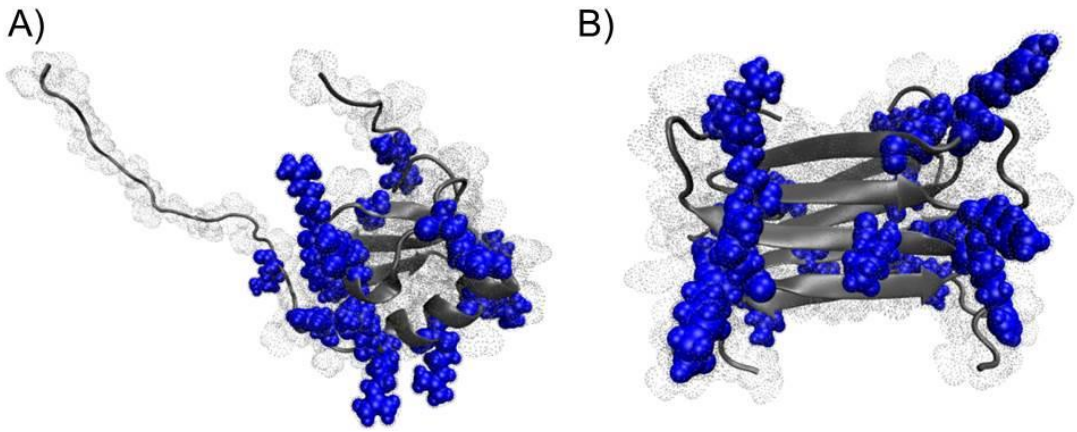

Figure 5.4: A) Ltn10 and B)Ltn40 with all basic residues labelled in blue with Van der Waals model, created in VMD from PDB files 1J9O and 2JP1 respectively.

The D_{WT}+Fx complex was also subjected to ECD fragmentation, however, only $[D_{WT}+Fx+11H]^{11+}$ and $[D_{WT}+Fx+12H]^{12+}$ were transmitted at high enough intensity to be examined here. At both charge states, however, no c- or z-type fragmentation was observed. A number of low intensity, low molecular weight fragments were detected and attributed to internal fragments occurring either from the dimer in a bound or unbound state; however, these fragments could not be assigned as any unique sequence and therefore cannot be unambiguously identified. The lack of significant, high intensity fragmentation demonstrates the stability of the complex and strength and specificity of binding. As with the monomeric bound form there is no observable loss of the ID tail in its entirety which is in contrast to the unbound dimer (Chapter 4 section 4.3.2), suggesting the ID tail is involved either in stabilising the complex or directly with binding. Furthermore, no significant fragmentation along the backbone occurs, either as a complex or as unbound dimer. In Ltn40 large basic

patches are again present on the surface of the protein which would provide a plausible region for Fx to bind (Figure 5.4), providing multiple interactions along the backbone explaining the lack of any high intensity fragmentation.

5.3.1.4 Assessing WT aggregation: transmission electron microscopy

As noted in section 5.3.1.1, WT Ltn samples aggregate upon addition of equimolar concentrations of Fx. TEM studies were then performed on these insoluble aggregates (Figure 5.5).

GAGs have been found to promote aggregation not only for chemokines but also for amyloid fibril-forming proteins, including α-synuclein[45] and β-amyloid protein[46,47]. Indeed GAGs are routinely found associated with, and incorporated within, amyloid deposits. Furthermore, heparin-induced aggregation of a number of the amyloid fibril forming proteins has been studied *in vitro* using TEM; however, to-date such studies have not been performed on chemokines. To better understand the aggregation process and species being formed such studies into chemokine:GAG aggregation is necessary.

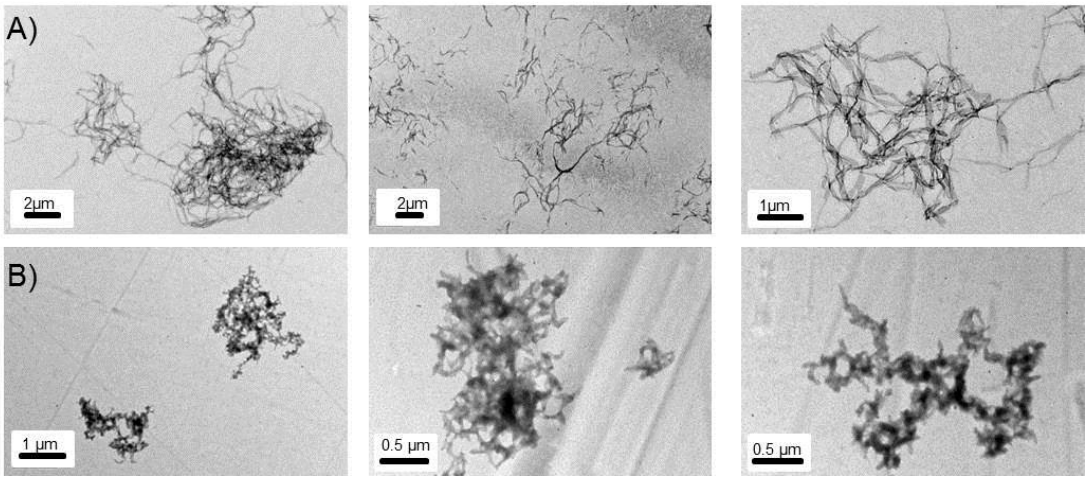

Figure 5.5: Representative TEM images obtained for 1:1 mixtures of Fx and A) WT Ltn (50 µM) B) WT Ltn (10 µM).

Considering WT Ltn in the presence of Fx at 50 µM protein concentration it is clear large aggregates are formed, which tend to form large clusters (Figure 5.5A). The morphology of the species formed is interesting, with ribbon-like structures appearing to twist around one and other. These observed species appear to have a degree of order or structure but are not consistent with the long straight protofilament structures frequently formed by amyloidgenic proteins, suggesting aggregation is different here. Heparin has, however, been observed to produce subtle changes in protein aggregation, promoting formation of flexible and curved

fibrils for α-synuclein[45], suggesting heparin can act as a template for fibril formation and hence its presence can also affect the morphology of the species formed.

Due to the overlapping nature of the species it is impossible to obtain an accurate length distribution under these conditions; visually species appear variable in length with a wide range of sizes typically observed in all images, normally on the order of μm. The size of both the individual structures and corresponding clusters of species is remarkable when considering samples were incubated for only 30 minutes prior to TEM grid preparation. When the concentration of the WT plus Fx sample is reduced (10 μM Ltn plus 10 μM Fx) (Figure 5.5B), the species formed are of a similar morphology, however, they appear shorter in length. The apparent shift in length distribution is ascribed to the lower concentration used here, limiting and slowing the aggregation process. It is clear, however, that even at low concentrations, for this type of *in vitro* study, significant aggregation occurs for WT Ltn upon addition of Fx.

A control study was undertaken to consider the individual solutions of WT Ltn and Fx (Appendix 4 Figure A4.3). For WT Ltn some small aggregates were present, whilst Fx exhibited limited signs of aggregation but of a different morphology than the species observed for Ltn plus Fx solutions, therefore it is important to note that Fx has a propensity for self-aggregation and hence is most likely promoting Ltn aggregation here.

5.3.2 Interactions with WT core and intrinsically disordered tail.

DT IM-MS and ECD studies into WT Ltn suggested that in all bound complexes the tail was most likely associated with or involved in binding, leading to tight conformations and no observable cleavage of the ID tail in ECD studies. In order to further probe this theory, the core of Ltn (WT 1-72) and the ID tail (WT 72-93) were then studied.

5.3.2.1 Ability of WT core and tail to bind Fx: MS studies

The WT 1-72 construct was first studied, this construct contains the full structural core of Ltn10 but does not contain the ID tail and therefore could be considered more similar to the traditional conserved chemokine structure than full length WT Ltn. Interestingly WT 1-72 aggregates much more extensively in the presence of Fx than the full length protein, with the spectra obtained at a 1:1 ratio at 50 μM protein concentration showing substantial aggregation, producing a spectrum of low quality (Appendix 4 Figure A4.4). In order to acquire a cleaner, clearer spectrum the sample was then prepared at a 2:1 (Ltn:Fx) ratio at a Ltn concentration of 25 μM (Figure 5.6B) and, as with WT Ltn, $M_{1\text{-}72}$+Fx, $D_{1\text{-}72}$+Fx and

$D_{1\text{-}72}$+2Fx species are formed, with observed and expected masses of all species being in good agreement (Appendix 4 Table A4.3). The considerable increase in aggregation observed for WT 1-72 in comparison with WT Ltn suggests that the presence of the ID tail inhibits GAG binding and aggregation, this could either be *via* direct interaction between the GAG and the tail or the tail may affect the specificity of binding and the corresponding complex conformations, with formation of aggregation-prone non-specific complexes in the absence of the tail.

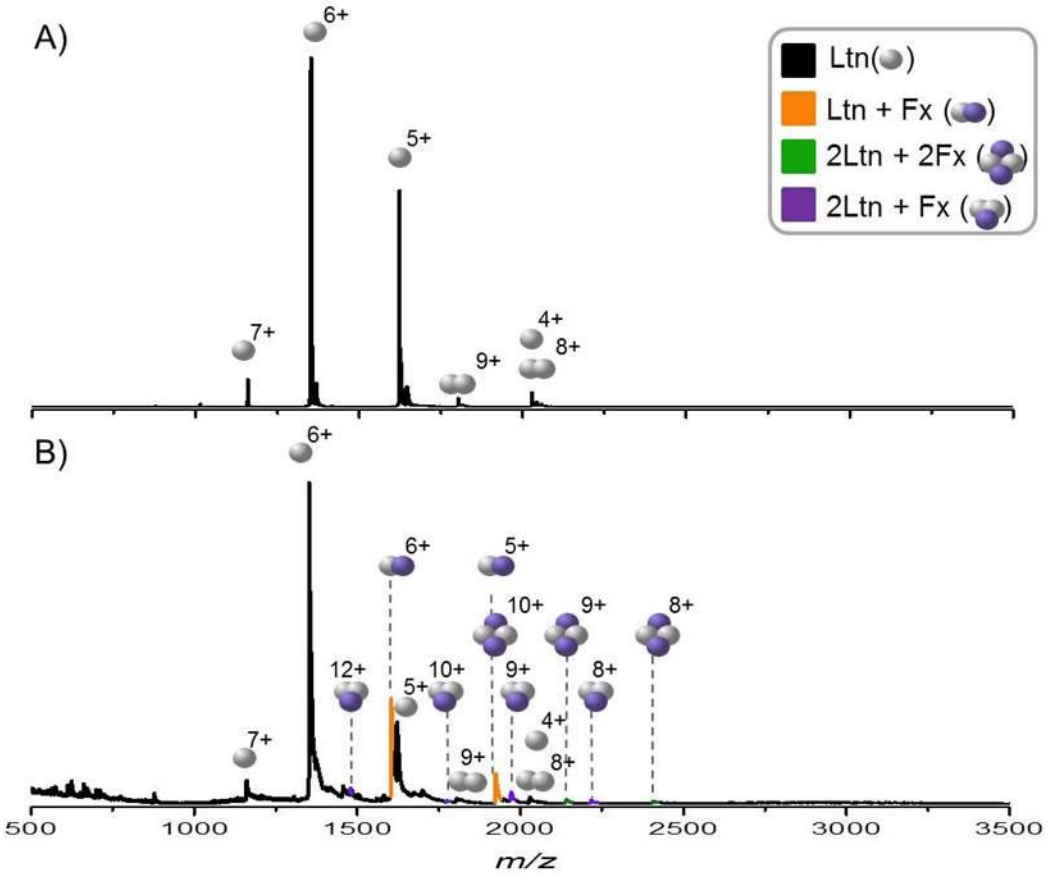

Figure 5.6: Spectra obtained for A) WT 1-72 (50 µM concentration) B) WT 1-72 plus Fx acquired at a 2:1 ratio at 25 µM Ltn concentration.

In order to further probe these interactions the ability of the ID tail (WT 72-93) to bind Fx was then considered. This peptide contains an additional two amino acids at its N-terminus as a product of his-tag cleavage and therefore does not fully represent the native tail but will indicate if this ID sequence of amino acids (which contains one basic amino acid, lysine in position 77) can bind Fx. The spectra obtained from a 1:1 mixture of WT 72-93 and Fx highlights that the tail itself is not responsible for GAG binding (Figure 5.7B) with no Fx-bound species observed by MS and no signs of extensive aggregation on these timescales

(~5 mins), TEM studies, however, show that aggregation can occur when incubated for 30 minutes (section 5.3.2.3). Therefore it is proposed instead that the ID tail affects the conformations of complexes adopted and the specificity of binding without directly interacting with Fx, a hypothesis which can be further probed by studying the conformations adopted by WT 1-72 upon binding to Fx.

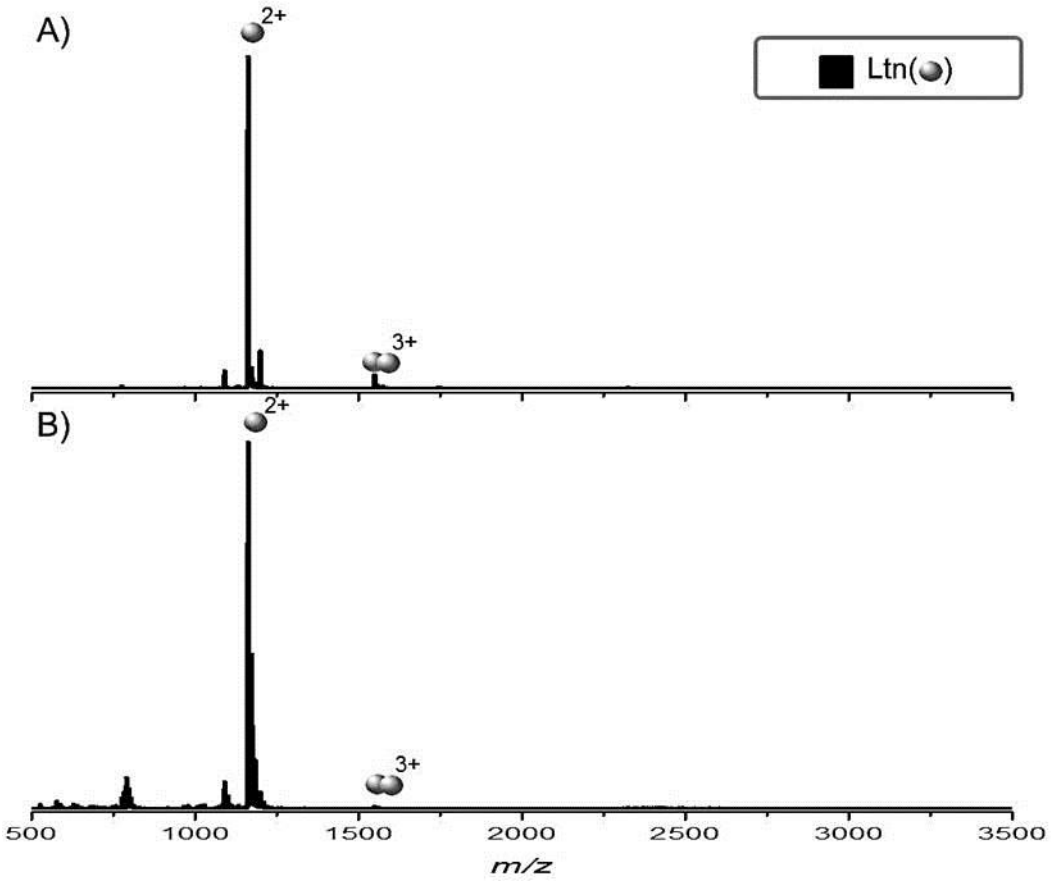

Figure 5.7: Spectra obtained for A) WT 72-93 B) WT 72-93 plus Fx prepared at 1:1 Ltn:Fx ratio. In both cases WT 72-93 was prepared at 50 μM concentration.

5.3.2.2 Conformations of WT 1-72:Fx complexes: DT IM-MS studies

Due to the inability of WT 72-93 to bind Fx, no DT IM-MS studies were performed on this construct. Instead, CCS of the WT 1-72:Fx species were determined. As with WT Ltn, comparisons were initially made between all observed unbound species in the absence and presence of Fx, Table 5.2.

The lower charge states of monomeric unbound WT 1-72 in the presence of Fx ($[M_{1\text{-}72}+5H]^{5+}$ and $[M_{1\text{-}72}+6H]^{6+}$) are found to be within experimental error of values determined in the absence of Fx. An additional, more extended $[M_{1\text{-}72}+6H]^{6+}$ species is, however, also identified in the presence of Fx, with a CCS of 1209 ± 13 Å^2. This more

extended species is of similar CCS to the extended conformation of $[M_{1-72}+7H]^{7+}$, suggesting it is most likely due to a species of the same conformational family and that the presence of Fx is causing WT 1-72 to unfold at lower charge states. Additionally, the more compact conformation of unbound $[M_{1-72}+7H]^{7+}$ is seen to increase in CCS in the presence of Fx, again consistent with Fx promoting unfolding. Fx clearly effects the conformations of unbound WT 1-72. There are a number of mechanisms through which this could occur, either Fx could be transiently, or non-specifically, interacting with WT 1-72 promoting unfolding of these species, alternatively, these species may be an encounter species *en route* to stable binding for which the initial interaction with Fx promotes extension and unfolding of the monomer following which stronger binding occurs forming the M_{1-72}+Fx species. Promotion of these extended conformations could explain the rapid aggregation observed by MS for this Ltn construct. No unbound dimeric WT 1-72 species are detected in the presence of Fx therefore the unbound species plotted in Figure 5.8 are those determined in the absence of Fx (Appendix 3 Table A3.2). The extended conformations identified for $[M_{1-72}+6H]^{6+}$ and $[M_{1-72}+7H]^{7+}$ in the presence of Fx are plotted in Figure 5.8.

Species	***CCS/ Å²***	
	Absence of Fx	***Presence of Fx***
$[M_{1-72}+5H]^{5+}$	916 ± 8	909 ± 13
$[M_{1-72}+6H]^{6+}$	950 ± 9	973 ± 20
$[M_{1-72}+6H]^{6+}$	--	1209 ± 13
$[M_{1-72}+7H]^{7+}$	1037*	1142 ± 18
$[M_{1-72}+7H]^{7+}$	1229 ± 12	1211 ± 3

Table 5.2: Experimentally determined average CCS for all unbound WT 1-72 species observed in the absence and presence of Fx, the error is reported as the standard deviation between repeats. *species observed in all repeats but only resolvable in one.

DT IM-MS findings for Fx-bound WT 1-72 (Figure 5.8 and Appendix 4 Table A4.4) tell a dramatically different story to the full length WT Ltn. For this mutant, multiple conformations, both for the M_{1-72}+Fx and the D_{1-72}+Fx species, are present. For species of both oligomeric orders CCS are larger than their unbound counterparts. Interestingly, the more extended conformation of $[M+Fx+6H]^{6+}$ is of similar CCS to that of the larger unbound 6+ species in the presence of Fx, suggesting the more extended unbound species may be *en route* to complex formation.

The observation of multiple conformations, which are often extended in nature, in combination with the extensive aggregation occurring during MS studies suggests that for this construct binding is less specific than for the full length WT Ltn. This suggests a possible structural role for the ID tail, despite this region showing no affinity to directly bind

Fx. It is plausible that the tail is needed to mediate specific binding and in doing so, also confers a conformational stability to the complexes.

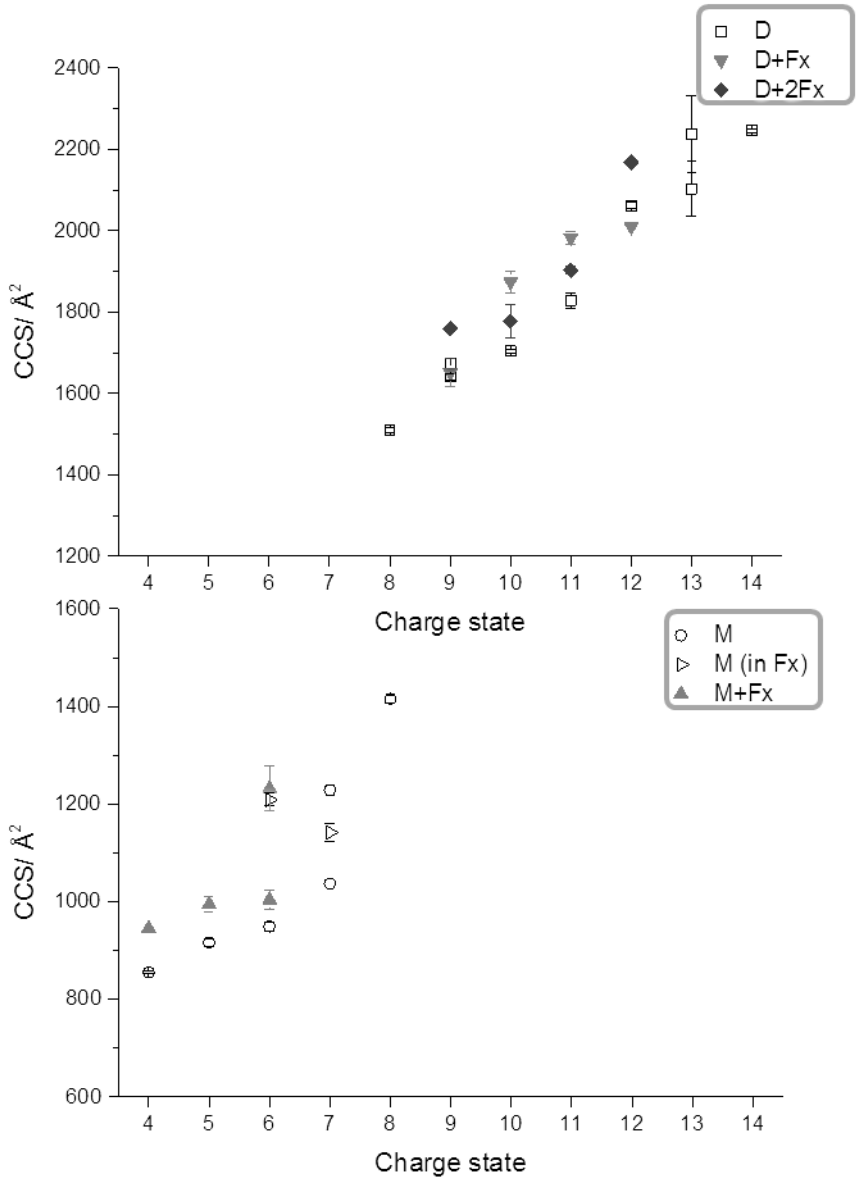

Figure 5.8: Average CCS of WT 1-72 Ltn and Fx bound Ltn dimeric (top) and monomeric (bottom) species, calculated from three repeats, error bars represent standard deviation between these values. Unbound species were determined in the absence of Fx (hollow symbols), extended species observed in the presence of Fx for $[M_{1\text{-}72}+6H]^{6+}$ and $[M_{1\text{-}72}+7H]^{7+}$ are represented by hollow right triangles. Filled symbols represent Fx bound species.

5.3.2.3 Assessing core and tail aggregation: transmission electron microscopy

MS and DT IM-MS results suggest that Fx binding and the corresponding aggregation process occurring for WT 1-72 are significantly different than that of WT Ltn, a finding further emphasised by the TEM images for this sample (Figure 5.9A). TEM images show larger globular aggregates forming, strikingly different in appearance to the ribbon-like species formed for WT Ltn. This finding supports the assertion that the ID tail is essential for formation of the structures observed for WT Ltn samples; it is conceivable that in WT Ltn the presence of this ID tail controls or limits the extent of aggregation. To-date no *in vitro* TEM studies into the aggregation process of chemokines in presence of GAGs have been reported. It would be interesting to compare the complexes formed by classical chemokines, in the presence of Fx, to Ltn, to determine if they demonstrate the ribbon or globular morphology, as there is a clear correlation between the two morphologies and the presence or absence of the ID tail for Ltn.

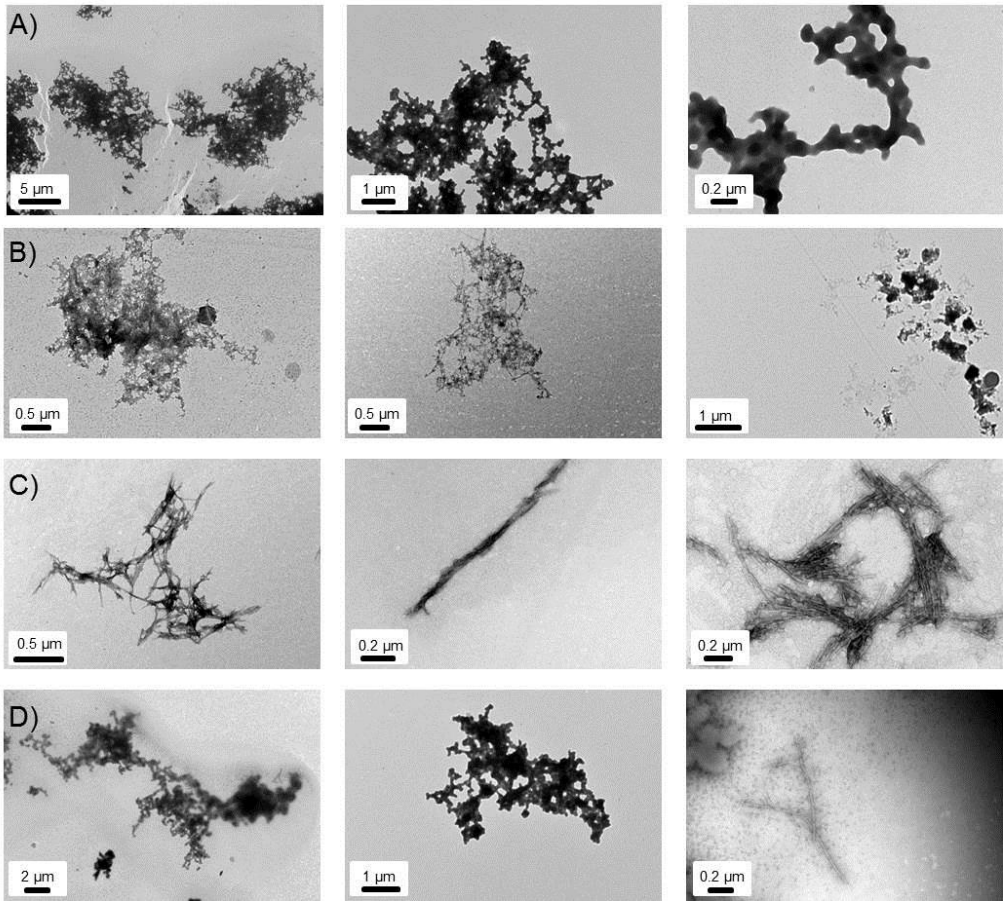

Figure 5.9: Representative TEM images obtained for A) WT 1-72 plus Fx (1:1 ratio at 50 µM) B) WT 72-93 (50 µM) C) WT 72-93 plus Fx (1:1 ratio at 50 µM) D) WT 1-72 plus WT 72-93 plus Fx (1:1:1 ratio at 50 µM).

In order to further study this effect, TEM images were acquired for the tail (WT 72-93) in the absence (Figure 5.9B) and presence (Figure 5.9C) of Fx. It is clear that the ID tail itself can aggregate forming clusters of interesting morphology, containing some straight species and overall appearing net-like in appearance. In the presence of Fx, however, this morphology changes, with primarily straight, short fibres which can align and twist around each other being observed. These appear more similar to the ribbon structures of the full length protein plus Fx than the species formed for WT 72-93 the absence of Fx, suggesting once again that this ID tail has a key role in the complexes formed for full length protein in the presence of Fx, despite no binding or aggregation being detected by MS. A WT 1-72 plus WT 72-93 plus Fx (1:1:1) solution was also studied (Figure 5.9D), with TEM images showing both the globular species formed for WT 1-72 plus Fx and the ribbon-like species observed for WT 72-93 plus Fx (Figure 5.9C). The TEM grid is, however, dominated by the globular species, illustrating that either these aggregates have a higher affinity for the grid or that aggregation of WT 1-72 occurs faster than WT 72-93 in the presence of Fx and hence the solution is dominated by these globular species. The latter is supported by further MS studies (Appendix 4 Figures 4.5 and 4.6) and suggests that non-covalent interactions between the tail peptide and the structural core, are not sufficient to inhibit this aggregation process. The combination of studies presented here illustrate that the ID tail plays a significant role in both the binding of Fx and the aggregation pathways of Ltn, proposing a functional role in GAG binding. These results could help explain why such a sequence would be preserved in nature.

5.3.3 Interactions of CC3 and Fx.

Due to the clear ability of WT Ltn to bind Fx as both a monomer and dimer the Ltn monomer mutant (CC3), which contains a second disulfide bond limiting its interconversion to Ltn40, was also studied to further probe these interactions.

5.3.3.1 Stoichiometry of binding: insights from MS

The spectrum for CC3 highlights that despite its conformational constriction it can bind Fx, mostly binding as a monomer; however, some low level dimeric bound species are also observed (Figure 5.10), with the observed and expected masses of all species being in good agreement (Appendix 4 Table A4.5). This observation is consistent with other chemokines, almost all of which contain two disulfide bonds and those studied have been shown to bind GAGs as monomers, dimers or even tetramers[2,36].

As with WT Ltn, dimeric CC3 species can bind either one or two Fx chains and their presence for this monomeric mutant is significant, suggesting that Fx is facilitating dimer formation. Whether dimeric CC3 is in the Ltn10, Ltn40 or a yet unknown conformation cannot be discerned here. The formation of dimeric CC3 Fx-bound species, has not been previously reported, most likely due to their low population. This work underlines the sensitivity of MS, for the study of these intricate and complex binding processes and highlights dimeric bound CC3 as an interesting target for future study to better understand how Fx is promoting dimerisation.

As with WT Ltn and WT 1-72 Ltn, CC3 appears to aggregate upon addition of Fx at equimolar concentrations, however, considering the baseline, signal intensity and visible aggregates observed in the tip it appears to aggregate to a lesser extent than WT Ltn.

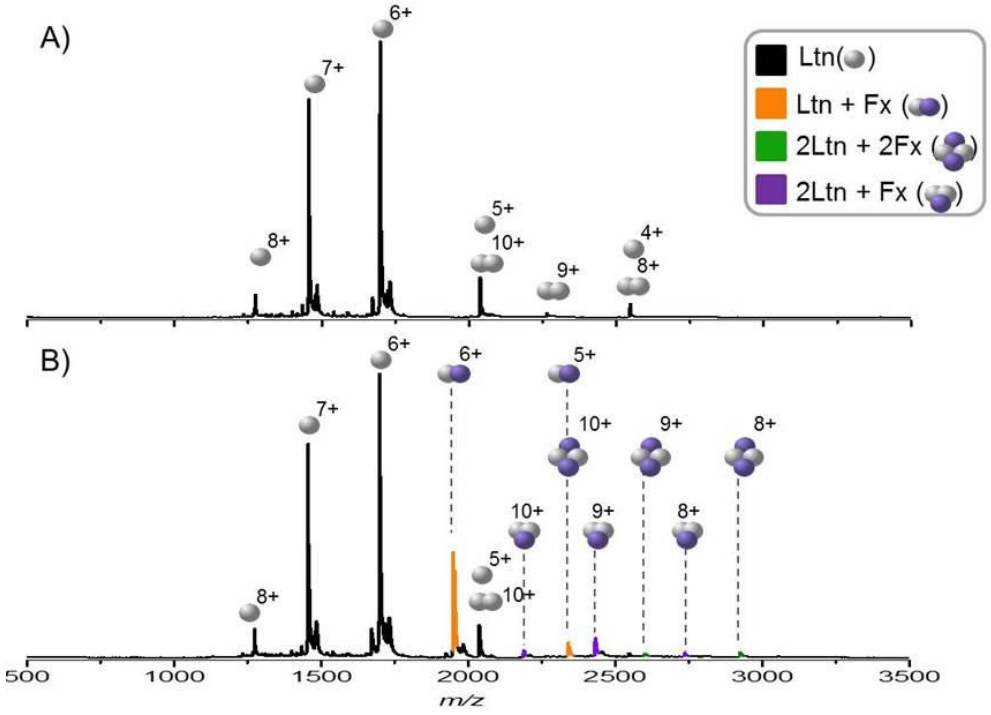

Figure 5.10 Spectra obtained for A) CC3 and B) CC3 plus Fx, 1:1 mixture. In both cases CC3 was prepared at 50 µM concentration.

5.3.3.2 Conformations of CC3:Fx complexes: DT IM-MS studies

For CC3, CCS of the two monomeric unbound species observed both in the presence and absence of Fx were found to be within experimental error (Table 5.3). This suggests that Fx does not significantly influence the conformations adopted by monomeric CC3. No unbound dimeric species were detected for CC3 in the presence of Fx due to their naturally low abundance so similar comparisons could not be made. Due to the low intensity and limited

charge state range of CC3 in the presence of Fx, on the DT IM-MS instrument, comparisons in Figure 5.11, were made with unbound CC3 in the absence of Fx (Appendix 3 Table A3.8).

Species	***CCS/ Å²***	
	Absence of Fx	***Presence of Fx***
$[M_{CC3}+6H]^{6+}$	1014 ± 3	1030 ± 28
$[M_{CC3}+7H]^{7+}$	1163 ± 4	1181 ± 23

Table 5.3: Experimentally determined average CCS for all unbound CC3 species observed in the absence and presence of Fx, the error is reported as the standard deviation between repeats.

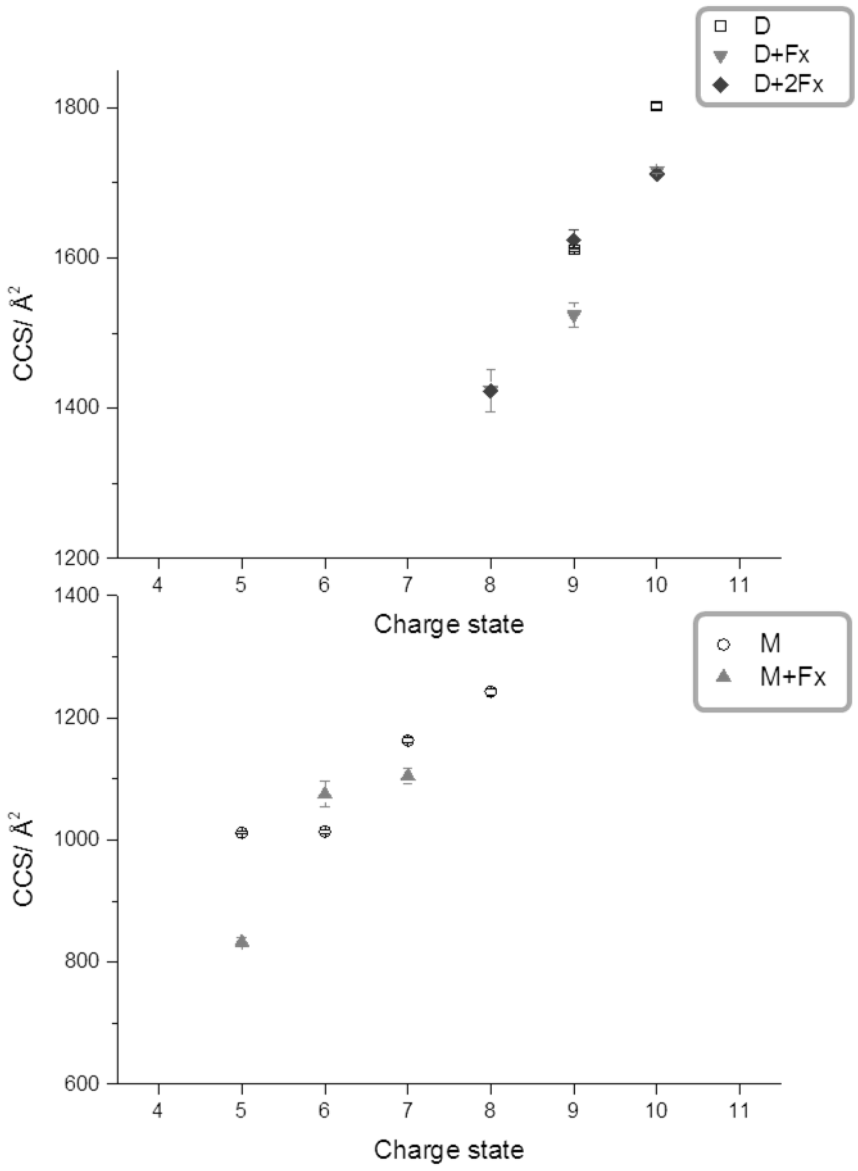

Figure 5.11: Average CCS of CC3 Ltn and Fx bound Ltn dimeric (top) and monomeric (bottom) species, calculated from three repeats, error bars represent standard deviation between these values. Unbound species were determined in the absence of Fx (hollow symbols). Filled symbols represent Fx bound species.

Conformations adopted by CC3 upon Fx binding were considered (Figure 5.11 and Appendix 4 Table A4.6). As with WT Ltn, the lowest charge state of the M_{CC3}+Fx is more

compact than unbound CC3 at the same charge state. The remaining two charge states ($[M_{CC3}+Fx+6H]^{6+}$ and $[M_{CC3}+Fx+7H]^{7+}$) are extremely similar in CCS, with $[M_{CC3}+Fx+6H]^{6+}$ being marginally larger than the unbound species at this charge state and $[M_{CC3}+Fx+7H]^{7+}$ being slightly more compact than its unbound counterpart, despite the additional mass due to Fx (1508 Da). These CCS suggest a stable, relatively compact conformational family is being formed upon Fx binding to monomeric CC3, a finding further supported by CID studies (Appendix 4 Figure A4.7). The variance in CCS for M_{CC3}+Fx is lower than that of the WT, changing only by ~30 % over the three charge states (as opposed to ~50% for M_{WT}+Fx), which is most likely due to the conformational constriction imposed on this mutant by the addition of the second disulfide bond.

The dimeric Fx bound species for CC3 follow a similar trend to that of WT Ltn, with both the D_{CC3}+Fx and D_{CC3}+2Fx presenting over a relatively narrow and similar range of CCS, with the D_{CC3}+2Fx species being slightly larger. The variance in CCS over the charge state range is again low ~20 % in each case, suggesting that these species are present in a stable conformational ensemble over which the addition of charge does not significantly affect the CCS. For both dimeric bound species it is interesting to note that CCS of the lowest charge states compare extremely favourably with the WT Fx-bound species of the same charge state, however as charge state is increased the CC3 species become larger than WT Ltn. The CCS obtained here for D_{CC3}+Fx and D_{CC3}+2Fx species suggests that these complexes are most likely stable and specific species, with Fx promoting the dimerization of this structurally constrained mutant.

5.3.3.3 Assessing CC3 aggregation: transmission electron microscopy

The aggregation occurring for CC3 in the presence of Fx was next considered (Figure 5.12), with species formed appearing visually extremely similar to those formed in the WT sample at the same concentration: large, twisted ribbon structures. CC3 plus Fx aggregation does not appear to form as numerous or extensive clusters as WT Ltn plus Fx suggesting that either these species have a significantly lower affinity for the TEM grid or CC3 aggregates to a lesser extent. The latter is consistent with observations made during, and findings from, MS studies which show evidence for lower aggregation levels; with a higher signal-to-noise and fewer visible aggregates observed in the n-ESI capillary. These smaller clusters formed for CC3 in the presence of Fx and viewed in the TEM images enable clearer and, therefore, more detailed images to be obtained on individual species and clearly show species twisting around each other.

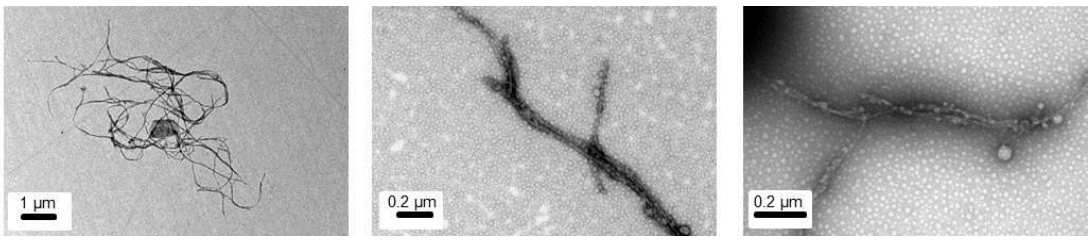

Figure 5.12: Representative TEM images obtained for CC3 plus Fx prepared at a 1:1 ratio at 50 µM concentration.

5.3.4 Interactions of W55D and Fx.

5.3.4.1 Stoichiometry of binding: insights from MS

The final Ltn mutant studied in detail here is W55D, which by MS is found to interact with Fx in a similar fashion to WT Ltn and CC3, forming M_{W55D}+Fx, D_{W55D}+Fx and D_{W55D}+2Fx species over a range of charge states (Figure 5.13), with observed and expected masses of all species being in good agreement (Appendix 4 Table A4.7). Fx-bound W55D species present over a wider charge state range than those of the WT Ltn or CC3 bound species, consistent with this highly flexible mutant presenting over a wider charge state and CCS range in its unbound form (see Chapter 4).

Interestingly, based on the relative intensity of the ions, W55D appears to bind Fx as a monomer to a lesser extent than WT Ltn, consistent with this mutant being mostly dimeric in solution; however, under these experimental conditions, in which extensive aggregation is observed, it is impossible to quantify the extent of binding. The fact that any Fx bound monomeric W55D is observed is significant, as NMR studies have shown this mutant to be mainly dimeric in solution and previous gas-phase structural studies into this mutant (Chapter 4 section 4.4.3) have suggested that monomeric signals in the spectra are due to dissociated dimer, occurring upon ionisation and desolvation. The occurrence of M_{W55D}+Fx species here, suggests either that there is a small proportion of monomer in solution which is capable of binding Fx or that the dimeric Fx-bound species, as with the unbound W55D dimer, are relatively fragile and can be broken apart upon ionisation and desolvation. CID experiments were performed on the highest intensity dimeric bound species, $[D_{W55D}+Fx+9H]^{9+}$, and showed dissociation to monomer and M_{W55D}+Fx, suggesting that M_{W55D}+Fx could indeed be a product of dissociation (Appendix 4 Figure A4.8). M_{W55D}+Fx were also subjected to CID fragmentation studies and similar observations were made to that of WT Ltn, with Fx fragmenting before complex dissociation; however, a greater proportion of monomer dissociation occurs for W55D (Appendix 4 Figure A4.9). Therefore it is not possible to discern the origin of these species in these initial binding studies and more in-

depth studies would need to be performed. Indeed it is possible that the M_{W55D}+Fx species observed are due to both possibilities described here, namely a small proportion of monomer being present in solution along with dissociation on desolvation.

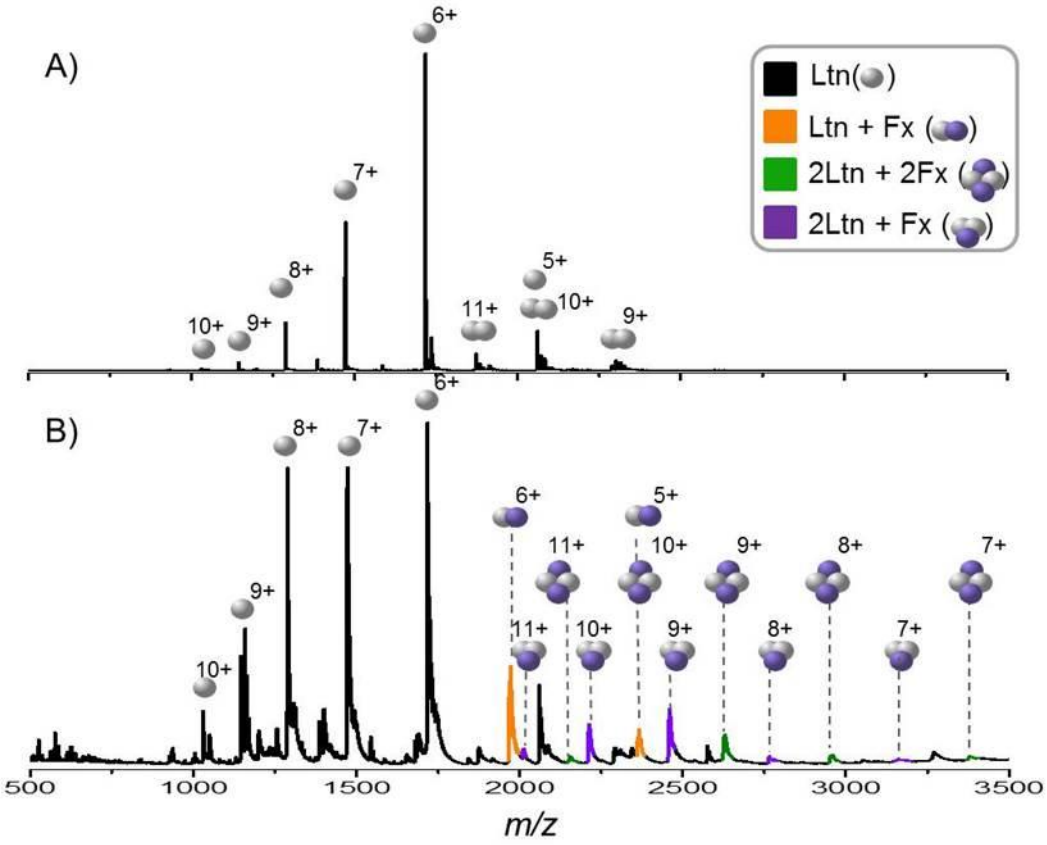

Figure 5.13: Spectra obtained for A) W55D and B) W55D plus Fx, 1:1 mixture. In both cases W55D was prepared at 50 µM concentration.

Comparison between the charge state distribution and intensity of the unbound W55D monomer in the absence (Figure 5.13A) and presence (Figure 5.13B) of Fx show clear differences. In the presence of Fx the signal intensity of the unbound species is greatly reduced, particularly for the species which accept either six or seven protons. This decrease in unbound monomer signal in the presence of Fx could be due to a number of factors: either Fx is binding a large proportion of dimeric protein in solution stabilising the dimer and limiting dissociation to monomer; the presence of Fx shifts the monomer:dimer equilibrium in solution reducing the intensity of any monomer present; or this unbound signal reduction could be due to W55D rapidly binding and aggregating, causing a significant decrease in all unbound protein signal in the presence of Fx. Indeed it is likely that this observation is due to a combination of these effects. It is clear that W55D is aggregating in the presence of Fx, due to the raised spectral baseline and observable aggregates in the n-ESI capillary, an

observation which is further probed using TEM (see section 5.3.4.3). It is, however, also possible that binding of Fx is stabilising dimeric W55D and limiting dissociation to monomer upon ionisation and desolvation.

5.3.4.2 Conformations of W55D:Fx complexes: DT IM-MS studies

For W55D CCS of all charge states of unbound dimeric species in the absence and presence of Fx are within experimental error, confirming that Fx does not affect the conformations of unbound dimeric W55D (Table 5.4).

The lowest charge states (4 and 5+) of unbound M_{W55D} were not observed in the presence of Fx and therefore in Figure 5.14 the CCS of these species are those determined in the absence of Fx. Unbound monomeric W55D species observed in the presence of Fx are, however, present in two conformational families, one of which is within experimental error to the species observed in the absence of Fx and whilst the other is more extended (Table 5.4). The more extended conformational family could be due to Fx affecting the conformations of any solution phase unbound monomeric species, promoting extension or unfolding. Alternatively, these more extended conformations could be due to dissociation of Fx-bound species during desolvation yielding monomeric species with a larger CCS. It is indeed likely that if dissociation of bound species was occurring upon desolvation that this would yield monomeric species of altered CCS. This observation coupled with significantly decreased monomeric signal in the presence of Fx (Figure 5.13) suggests that Fx binding is stabilising the W55D dimer, leading to decreased dissociation upon ionisation and desolvation, hence a lower unbound signal. Furthermore, when dissociation does occur it yields monomers of different CCS due to the structure of the precursor complex. It is therefore speculated that monomeric species of extended CCS are due to dissociation from dimeric W55D:Fx complexes.

Species	***CCS/ Å²***	
	Absence of Fx	***Presence of Fx***
$[M_{W55D}+6H]^{6+}$	1057 ±12	1075 ± 20
$[M_{W55D}+6H]^{6+}$	--	1238
$[M_{W55D}+7H]^{7+}$	1150 ± 26	1165 ± 34
$[M_{W55D}+7H]^{7+}$	--	1480 ± 26
$[M_{W55D}+8H]^{8+}$	1401 ± 6	1408
$[M_{W55D}+8H]^{8+}$	--	1717
$[D_{W55D}+10H]^{10+}$	1704 ± 6	1687
$[D_{W55D}+11H]^{11+}$	2020	1998
$[D_{W55D}+12H]^{12+}$	2062 ± 8	2059
$[D_{W55D}+13H]^{13+}$	2103 ± 67	2171

Table 5.4: Experimentally determined average CCS for all unbound W55D species determined in the absence and presence of Fx, the error is reported as the standard deviation between repeats.

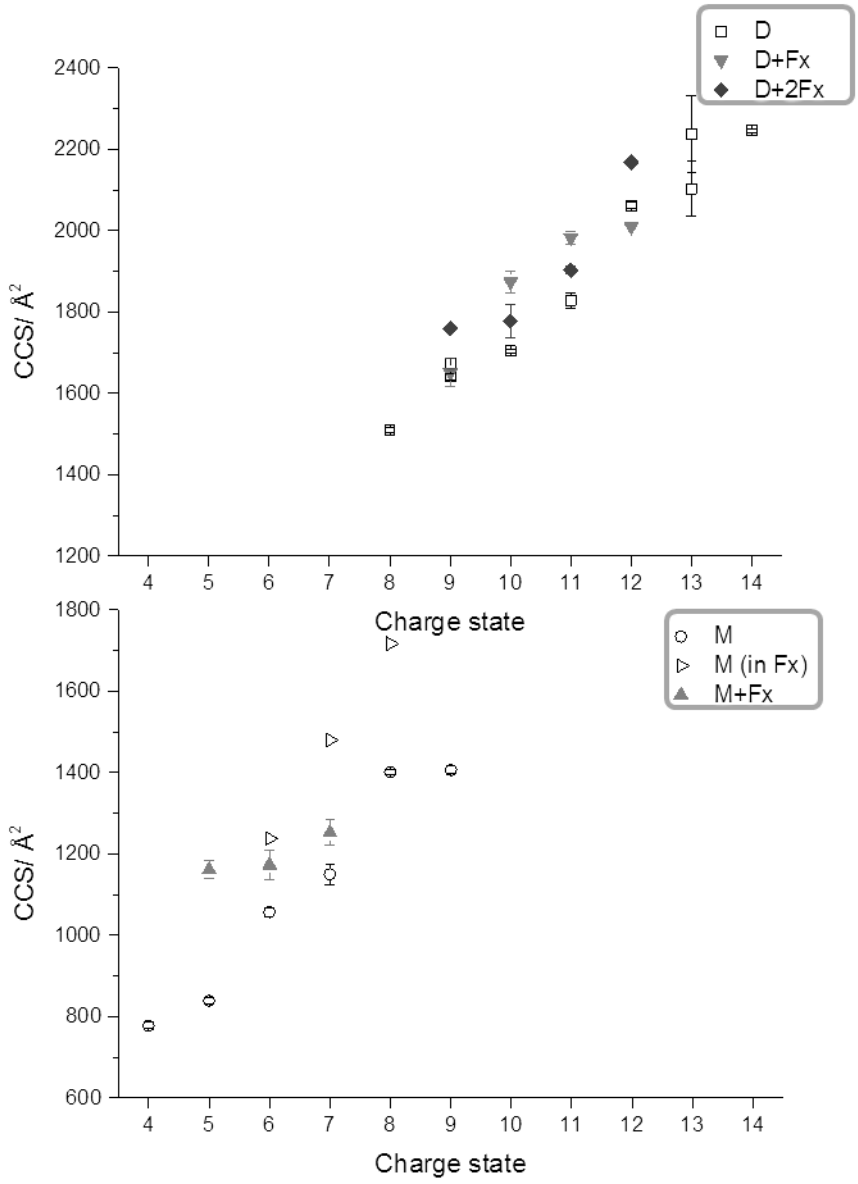

Figure 5.14: Average CCS of W55D and Fx bound W55D dimeric (top) and monomeric (bottom) species, calculated from three repeats, error bars represent standard deviation between these values. Unbound species were determined in the absence of Fx (hollow symbols), extended species observed in the presence of Fx for $[M_{W55D}+6H]^{6+}$ to $[M_{W55D}+8H]^{8+}$ are represented by hollow right triangles. Filled symbols represent Fx bound species.

Considering the W55D:Fx complexes (Appendix 4 Table A4.8), CCS determined for monomeric bound species of W55D (Figure 5.14) lie between the two conformational families of unbound monomeric W55D. Interestingly these M_{W55D}+Fx species present over a much narrower range of CCS than determined for M+Fx of both WT and CC3 Ltn, increasing by only ~10 % over the charge state range. This finding is in contrast to the trends for the unbound species, which highlighted that W55D in both its monomeric and dimeric form is more flexible than both WT and CC3 Ltn. Binding of Fx thus has a constricting

effect on the monomeric form of W55D causing a decrease in flexibility. This is consistent with the monomeric form of W55D being different to that of the conserved chemokine fold, as highlighted in Chapter 4 Section 4.4.3, and suggests that the M_{W55D} and M_{W55D}+Fx species could be a product of dissociation, possibly from a dimeric bound form. In-depth studies outside of the scope of this initial binding study would, however, be needed to determine the likely conformations M_{W55D}+Fx has adopted. Even without knowing the precise conformation these species are adopting, useful comparisons are made between the different constructs and between monomeric and dimeric bound species. This observation also strongly reinforces a point raised from Chapter 4, that a more stable mutant is required to properly mimic the Ltn40 fold.

D_{W55D}+Fx present with CCS both smaller (charge states 9+ and 12+) and larger (charge states 10+ and 11+) than the unbound W55D dimer, however, CCS increase by only ~20 % over this charge state range, suggesting a stable conformational family is formed. These W55D bound dimers are significantly larger than the WT and CC3 D+Fx species, consistent with unbound W55D generally being larger.

D_{W55D}+2Fx species present over a similar CCS range as the D_{W55D}+Fx species, with conformations both smaller and larger than D_{W55D}+Fx at the same charge state and larger than the unbound dimeric species at all charge states. CCS of D_{W55D}+2Fx increase by only ~20 % over the charge state range, again consistent with a stable conformational family of the Fx bound form and highlighting the specificity and stabilising effect of Fx binding. It is therefore surmised that as for WT and CC3 Ltn, W55D forms specific stable dimeric Fx bound complexes.

5.3.4.3 Assessing W55D aggregation: transmission electron microscopy

The species formed in the W55D plus Fx mixed sample (Figure 5.15) appear by TEM much shorter in length than those formed for WT Ltn and CC3 at the same concentration, although the morphology appears similar, forming ribbon-like species. It is notable that a single point mutation can cause such a difference in the length of these large aggregate species.

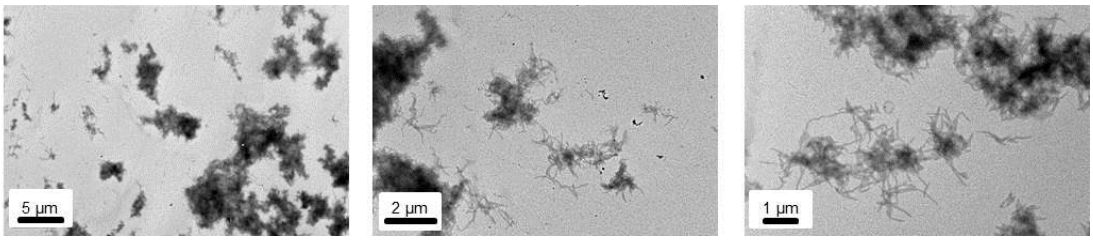

Figure 5.15: Representative TEM images obtained for W55D plus Fx prepared at a 1:1 ratio at 50 µM concentration.

The single point mutation in W55D is thought to push the solution phase equilibrium towards the dimeric structure, whereas WT and CC3 have a substantial proportion of monomer in solution. DT IM-MS experiments of WT and CC3 show that the M+Fx species are highly flexible in nature, with a wide range of CCS determined. It is therefore feasible to speculate that in the case of CC3 and WT Ltn, the flexibility of these M+Fx bound species may lead to the large twisted ribbon-like species formed for these two constructs. In contrast the W55D with a lower proportion of monomer, if any, in solution leads to less flexible M_{W55D}+Fx species being formed, potentially from dimeric bound structures in solution, which due to their rigidity cannot form the same long, flexible aggregates observed for WT and CC3, consistent with the decreased flexibility of the M_{W55D}+Fx species. If this is the case it would suggest that, in this set of experiments, the species detected by DT IM-MS could be on-pathway to the large aggregate species in the TEM studies, especially when considering the short time scales of these experiments, and hence such approaches could be useful in future in-depth analysis of this aggregation mechanism and pathway.

5.3.5 Probing the GAG binding function through further specific site mutations

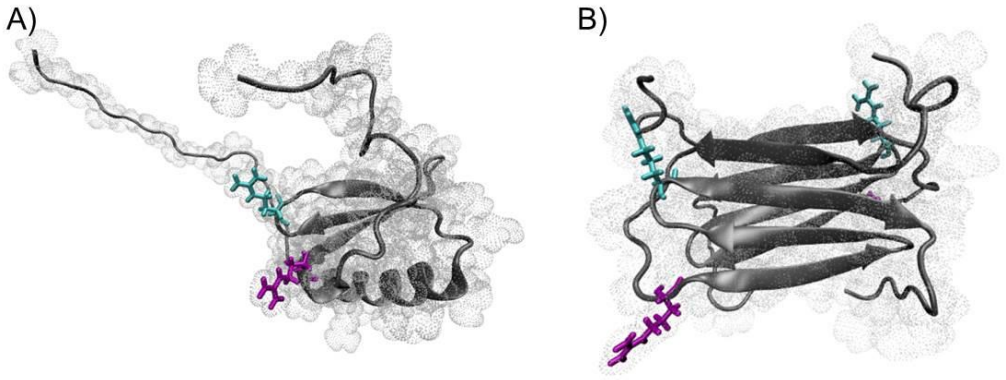

Figure 5.16: Arginine residues 23 and 43 labelled in purple and cyan respectively, on A) Ltn10 and B) Ltn40, both chains labelled. Figure created in VMD from PDB files 1J9O and 2JP1 respectively

The Volkman group[41] have identified that residues R23 and R43 are key for GAG binding, with binding affected by a factor of 40, as measured in surface plasmon resonance (SPR) sensograms for R23A and R43A compared to WT Ltn. Furthermore, the same study highlighted that the R23A/R43A double point mutant showed a further decrease in binding, by a factor of 300. The position of these residues are shown mapped on to the Ltn10 (Figure 5.16A) and Ltn40 folds (Figure 5.16B). This study was performed prior to the Ltn40 conformation being solved and, therefore, solely considered the Ltn10 fold. The point mutants do, however, show a significant decrease in GAG binding both by SPR and

heparin-sepharose column binding studies, during which the proteins eluted in the following order: WT, R23A, R43A and R23A/R43A (Table 5.5). Therefore in order to further probe the interactions between Fx and Ltn, preliminary MS and TEM studies were performed on these mutants of reduced affinity. The spectra of all three arginine mutants can be found in Appendix 4 Figure 4.10.

Ltn mutant	*Peak 1 elution (mM NaCl)*	*Peak 2 elution (mM NaCl)*
WT	452	694
R23A	382	557
R43A	376	458
R23A/R43A	309	344

Table 5.5: Elution from heparin-sepharose column for WT, R23A, R43A and R23A/R43A[39]. Peak 1 is assigned as Ltn10 and peak 2 as Ltn40 based on studies of WT, CC3 and W55D.

5.3.5.1 Insights in to interactions: MS studies.

R23A Ltn was the first of this series of mutants to be studied, as by heparin-sepharose column binding studies this mutation was shown to have the smallest effect on heparin binding both in its monomeric (peak 1) and dimeric (peak 2) form, see Table 5.5. Compared to the mass spectra acquired for WT Ltn, in the presence of Fx under similar conditions (Figure 5.17A) R23A appears to favour a dimeric Fx-bound form (Figure 5.17B), with these species having a higher relative intensity than for WT Ltn. A significant observation considering this mutation has been shown to shift the unbound equilibrium slightly towards the monomeric fold[39]. Direct comparisons between these spectra, however, cannot be made as these proteins could be, and most likely are, aggregating to different extents which would affect the spectra acquired and intensity of all species within the spectra. Visually, it does appear as if R23A is aggregating to a lesser extent than WT Ltn in the presence of Fx; however, TEM studies enable this to be probed in detail (see section 5.3.5.2). From the spectra obtained it is clear that R23A is capable of binding Fx, and can bind in both a monomeric and dimeric form and hence this mutation does not completely diminish GAG binding.

Considering next the R43A Ltn mutant, the spectrum obtained (Figure 5.17C) shows a decrease in intensity of dimeric Ltn binding Fx with respect to R23A. This observation is consistent with this mutant binding GAGs with lower affinity than R23A and WT Ltn, hence solidifying that this residue is of importance with respect to GAG binding. Furthermore, R43A appears to aggregate to a lesser extent than previous Ltn constructs studied, indicative of a lower GAG binding affinity. As with R23A, R43A can, however, still bind Fx in both a monomeric and dimeric form, highlighting that although R43 is clearly important for GAG

binding, mutation of this residue to an alanine does not completely diminish the Ltn GAG binding propensity.

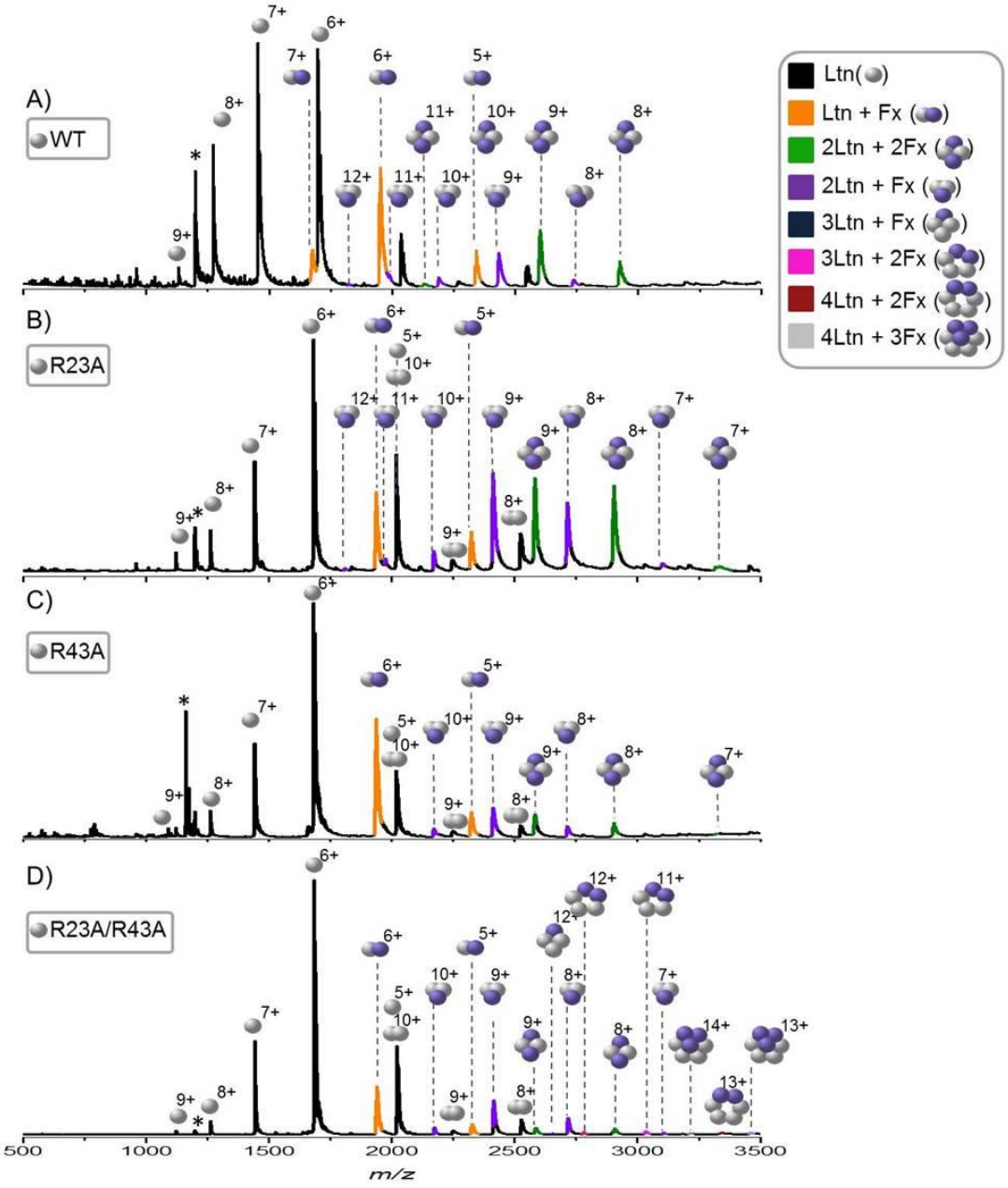

Figure 5.17: Spectra obtained for Ltn plus Fx for A)WT, B) R23A , C) R43A and D) R23A/R43A. In all cases spectra were obtained from 1:1 mixtures of Ltn:Fx at 50 μM protein concetration. Expected and observed masses of all species can be found in Appendix 4 Table A4.9.

The final arginine/alanine mutant studied was the double point mutation R23A/R43A which was shown by both SPR and heparin-sepharose column binding studies to have the most significant decrease in binding for both the monomeric and dimeric forms. Mutation of either and more significantly both of these amino acids would highly disrupt the basic patches of

amino acids present on both folds of Ltn (Figure 5.4), reducing the GAG binding affinity of both monomeric and dimeric Ltn. The spectra obtained for R23A/R43A plus Fx (Figure 5.17D) shows a reduction in both monomeric and dimeric binding of Ltn, with these species present at a much lower intensity, consistent with heparin-sepharose column binding studies which revealed a significant decrease in Ltn10 and Ltn40 binding for R23A/R43A. The R23A/R43A construct was also visually observed to aggregate to a lesser extent than previous samples studied, such as WT Ltn.

Intriguingly, a number of additional species are present in this spectrum, assigned as Fx-bound trimers and tetramers of R23A/R43A (Figure 5.18). The expected and observed masses of all species for R23A/R43A, and the two single point mutants, are in good agreement (Appendix 4 Table A4.9). These higher order oligomers were not observed for any other Ltn construct studied here and could be unique, favoured complexes for R23A/R43A, or could be attributable to slower aggregation. For this system it may be feasible to capture oligomeric species on pathway to higher order aggregates. If this were the case, however, it would be expected that a number of species of increasing oligomeric orders would have been observed for the lower concentration WT samples or for the other Ltn arginine mutants which aggregate to a lesser extent. Therefore, it is thought that the identification of these species is not due to the slower aggregation occurring here and instead formation of these species is specifically promoted *via* the mutations made to R23A/R43A Ltn.

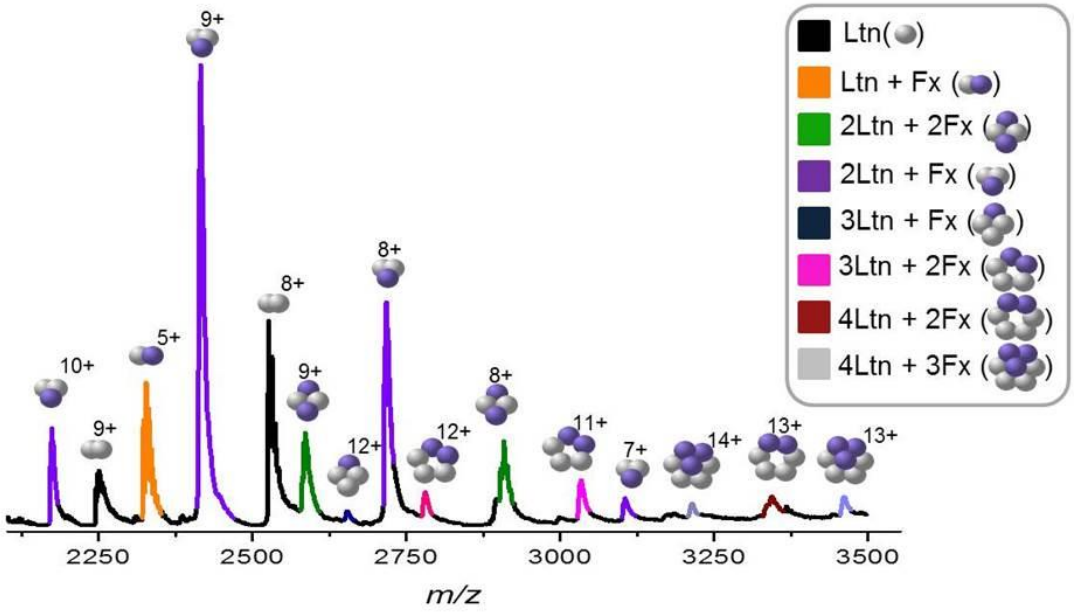

Figure 5.18: Enlarged region of *m/z* 2100-3500 for spectrum shown in Figure 5.17E for R23A/R43A plus Fx prepared at a 1:1 ratio at 50 µM protein concentration.

R23A/R43A Ltn does not form trimers or tetramers in the absence of Fx (Appendix 4 Figure A4.10). Furthermore, no unbound trimers or tetramers are observed in the presence of Fx, emphasising that formation of these species is induced by Fx to enable binding to occur. The chemokines CCL2 and CXCL4 have been shown to bind GAGs as tetramers, with detailed structural studies highlighting that tetramerisation of these chemokines results in a structure containing continuous basic patches providing a GAG binding surface[18,21]. This provides a possible explanation for the observation of Fx bound trimers and tetramers for R23A/R43A. These two residues have proved key for the GAG binding of WT Ltn and fall in the centre of the large basic patch of amino acids on both Ltn10 and Ltn40, however, WT Ltn contains a further 13 basic residues; seven arginine residues and six lysine residues. It is therefore speculated here that to overcome the detrimental effect, with respect to GAG binding, of removing two basic residues R23A/R43A oligomerises, with the packing of the subunits creating a basic surface to which the GAG can bind. Here, it is thought that oligomerisation occurs prior to stable R23A/R43A:Fx complex formation, most likely through transient interactions with Fx, after which stable species can form. DT IM-MS in combination with detailed MD studies has the potential to shed light on the conformations of these species and further work should involve such studies. The intensity of all Fx-bound species is low, accentuating the fact that GAG binding affinity is significantly reduced in this construct and this cannot be completely overcome through the rapid formation of higher order oligomers.

5.3.5.2 Assessing aggregation: TEM studies

The three arginine mutants were also studied using TEM, to assess the extent to which they were aggregating in the presence of Fx and if the morphology of the species was similar to that observed for the other Ltn constructs studied here.

As with all other constructs, samples were allowed to incubate for 30 minutes prior to TEM grid preparation. R23A forms aligned structures of a ribbon-like morphology (Figure 5.19A), similar to those for WT (Figure 5.5), CC3 (Figure 5.12) and W55D (Figure 5.15) Ltn constructs and the Ltn ID tail (Figure 5.9B), in the presence of Fx. The sample, however, is not observed to aggregate to the same extent as the other Ltn mutants studied here and reiterates that GAG binding and hence GAG induced aggregation is reduced for this mutant. R43A Ltn produces ribbon-like structures (Figure 5.19B), shorter in length than those of WT and CC3 Ltn plus Fx, attributed to the lower GAG binding affinity and hence slower GAG induced aggregated for this mutant. It is interesting, however, to note that both the R23A and R43A single point mutations do not significantly alter the morphology of the large aggregates, but do limit the extent of aggregation.

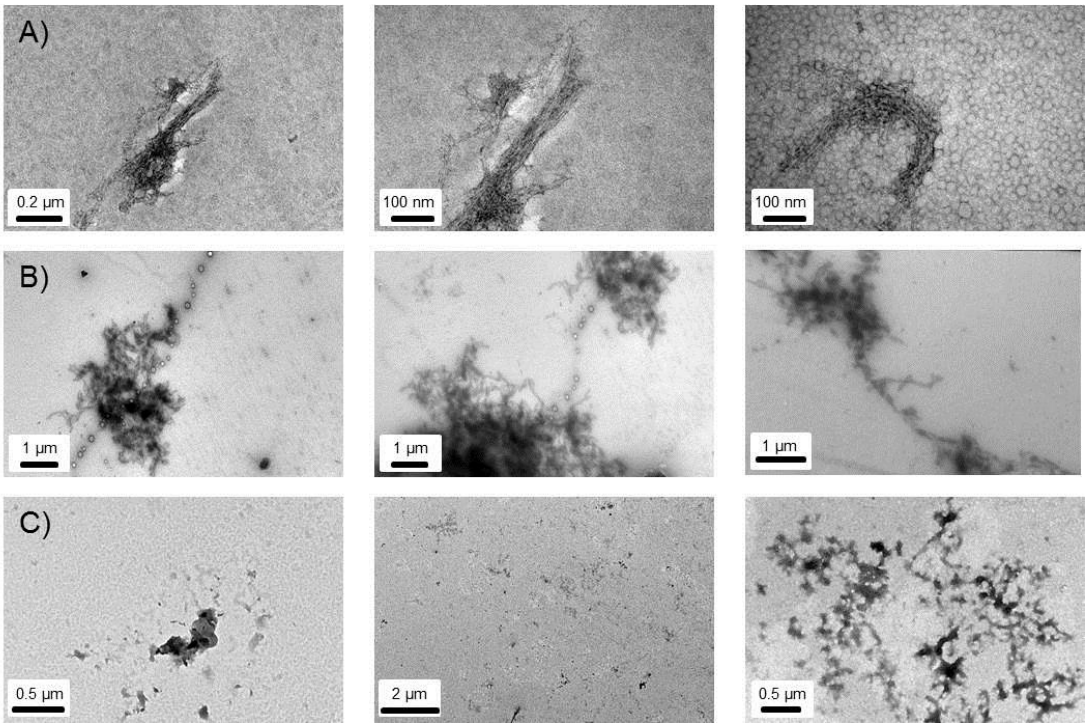

Figure 5.19: Representative TEM images obtained for 1:1 mixtures of Fx and A) R23A B) R43A C) R23A/R43A.

The R23A/R43A mutant displays very little aggregation, with the majority of images showing either dye aggregation or low level protein aggregation (Figure 5.19C). The aggregates observed for R23A/R43A appear less structurally defined than other Ltn species, and form much smaller clusters consistent with this mutant binding Fx with the lowest affinity and undergoing minor GAG induced aggregation.

5.4 Conclusions

The studies presented here focused on one of the functions of the metamorphic chemokine Ltn, namely to bind glycosaminoglycans, using a combination of MS-based techniques to provide details on stoichiometry of binding and conformations adopted by bound species, in addition to TEM studies enabling the visualisation of aggregation occurring upon the addition of GAGs.

In the main this work focused on the interactions of WT Ltn, mutants designed to mimic the Ltn10 (CC3) or Ltn40 (W55D) folds in addition to the ability of the structural core (WT 1-72) and ID tail (WT 73-93) to bind a model pentameric GAG, fondaparinux (Fx). The

results highlight that, despite its metamorphic nature, the functions of WT Ltn are not completely separate and defined *in vitro*, and Ltn is capable of binding Fx as both a monomer and a dimer. WT Ltn forms monomeric complexes capable of binding one Fx chain and dimeric complexes capable of binding one or two Fx chains, with DT IM-MS experiments indicating stable, specific complexes are being formed in all cases. The monomeric bound species are inherently flexible and present over a wide range of CCS (increasing in CCS by up to 50 %); this flexibility is thought, in part, to enable the formation of large flexible, twisted ribbon-like structures to be formed during the GAG induced aggregation process.

Of all Ltn constructs studied in detail, CC3 appears most similar to WT Ltn, with respect to stoichiometry of binding, conformations of monomeric and dimeric bound species and significantly the larger aggregates visualised through TEM studies, despite the more conformationally confined nature of this construct. Even after pushing the equilibrium strongly towards the monomeric form of Ltn (through addition of an extra disulfide bond) the nature of GAG binding is not perturbed.

The dimeric mutant W55D provides more complex results and points to its inadequacy as a dimeric mutant. Binding of Fx does reduce the conformational flexibility of dimeric W55D, with Fx bound species presenting over a much narrower range of CCS than its unbound counterparts. Monomeric bound species of W55D also appear more structurally constrained than both the unbound monomer and the other constructs studied here, which infers that these species are of a different nature than the other mutants studied, consistent with the earlier structural studies of W55D (Chapter 4). Interestingly, by TEM W55D is observed to form aggregates of a similar morphology but of much shorter length than WT and CC3 in the presence of Fx, an observation which has been attributed to the decreased flexibility of the W55D bound species.

Significantly, WT 1-72 has substantially different behaviour in the presence of Fx, with extensive aggregation occurring, forming species of an altered, globular morphology. In addition the GAG-bound complexes adopt multiple, highly extended conformations suggesting binding is less specific for this mutant. The ID tail (WT 72-93) itself does not bind Fx (as viewed by MS), however, based on the results for WT 1-72 in comparison to all full length constructs, the ID tail is essential for specific, stable GAG-bound conformers to be formed, proposing a structural role for this region.

In addition a series of arginine mutants (R23A, R43A and R23A/R43A) which had been previously reported to exhibit decreased GAG binding[39,41] were studied. For all three mutants, a decrease in intensity of binding and GAG-induced aggregation is observed. For the single point mutants this effect is more pronounced for R43A, than for R23A, indicating that the R43 residue is more significant with respect to GAG binding and particularly for dimeric GAG binding. Both the R23 and R43 residues lie within a large basic patch, in both the monomeric and dimeric folds and hence mutation of either of these residues disrupts this basic region, which then inhibits binding. Mutation of both the R23 and R43 residues to form R23A/R43A reduces the intensity of GAG binding further attributed to greater disruption of the basic region. R23A/R43A, however, shows by MS, higher order oligomers binding Fx, with trimers and tertramers being detected. These species were not present in the spectra of the protein itself or in any other Ltn mutant spectra, in the absence or presence of Fx. It is therefore hypothesised that R23A/R43A oligomerises forming trimers and tetramers in order to create a larger basic patch of amino acids to which Fx can then bind, in order to overcome the disruption to this region caused *via* the mutations. It is speculated here that weak, transient interactions between R23A/R43A and Fx promote oligomerisation of R23A/R43A enabling stronger, more specific interactions between the oligomers and Fx to occur. Future studies into these mutants should involve DT IM-MS experiments to obtain information on the conformations of these species.

Clearly, GAG binding is a complicated and intricate process; however, the *in vitro* work presented here probes this in detail. It is surmised that Fx binds to a basic patch of amino acids on both the Ltn10 and Ltn40 folds. Interactions between GAGs and a basic region of a protein would enable multiple interactions to occur, strengthening and stabilising binding and corresponding complexes formed. Mutation of basic amino acids within this accessible region decreases the GAG binding propensity of Ltn. Furthermore, complexes formed between Ltn and Fx appear to be stabilised by the ID tail, suggesting this preserved disordered region is functionaly important with respect to GAG binding, despite not being needed for receptor activation[37]. Finally, this combination of biophysical tools clearly can be fruitful in the study of chemokine:GAG binding.

5.5 References

(1) Handel, T.; Hamel, D. *Chemokines*; Elsevier: San Diego, 2009.
(2) Handel, T. M.; Johnson, Z.; Crown, S. E.; Lau, E. K.; Proudfoot, A. E., *Annu. Rev. Biochem.*, **2005**, *74*, 385-410.
(3) Shriver, Z.; Liu, D.; Sasisekharan, R., *Trends Cardovas. Med.*, **2002**, *12*, 71-77.
(4) Lindahl, U.; Hook, M., *Annu. Rev. Biochem.*, **1978**, *47*, 385-417.
(5) Ihrcke, N. S.; Wrenshall, L. E.; Lindman, B. J.; Platt, J. L., *Immunol. Today*, **1993**, *14*, 500-505.
(6) Johnson, Z.; Proudfoot, A.; Handel, T., *Cytokine Growth Factor Rev.*, **2005**, *16*, 625-636.
(7) Murphy, P. M., *Annu. Rev. Immunol.*, **1994**, *12*, 593-633.
(8) Proudfoot, A. E. I.; Handel, T. M.; Johnson, Z.; Lau, E. K.; LiWang, P.; Clark-Lewis, I.; Borlat, F.; Wells, T. N. C.; Kosco-Vilbois, M. H., *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, *100*, 1885-1890.
(9) Webb, L.; Ehrengruber, M. U.; Clark-Lewis, I.; Baggiolini, M.; Rot, A., *Proc. Natl. Acad. Sci. U. S. A.*, **1993**, *90*, 7158-7162.
(10) Witt, D. P.; Lander, A. D., *Curr. Biol.*, **1994**, *4*, 394-400.
(11) Hoogewerf, A. J.; Kuschert, G. S.; Proudfoot, A. E.; Borlat, F.; Clark-Lewis, I.; Power, C. A.; Wells, T. N., *Biochemistry*, **1997**, *36*, 13570-13578.
(12) Lalani, A. S.; Graham, K.; Mossman, K.; Rajarathnam, K.; Clark-Lewis, I.; Kelvin, D.; McFadden, G., *J. Virol.*, **1997**, *71*, 4356-4363.
(13) Alexander, J. M.; Nelson, C. A.; van Berkel, V.; Lau, E. K.; Studts, J. M.; Brett, T. J.; Speck, S. H.; Handel, T. M.; Virgin, H. W.; Fremont, D. H., *Cell*, **2002**, *111*, 343-356.
(14) van Berkel, V.; Barrett, J.; Tiffany, H. L.; Fremont, D. H.; Murphy, P. M.; McFadden, G.; Speck, S. H.; Virgin, H. W., *J. Virol.*, **2000**, *74*, 6741-6747.
(15) Ruiz-Argüello, M. B.; Smith, V. P.; Campanella, G. S.; Baleux, F.; Arenzana-Seisdedos, F.; Luster, A. D.; Alcami, A., *J. Virol.*, **2008**, *82*, 917-926.
(16) Hamel, D. J.; Sielaff, I.; Proudfoot, A. E.; Handel, T. M., *Methods Enzymol.*, **2009**, *461*, 71-102.
(17) Kuschert, G. S.; Hoogewerf, A. J.; Proudfoot, A. E.; Chung, C.-w.; Cooke, R. M.; Hubbard, R. E.; Wells, T. N.; Sanderson, P. N., *Biochemistry*, **1998**, *37*, 11193-11201.
(18) Stringer, S. E.; Gallagher, J. T., *J. Biol. Chem.*, **1997**, *272*, 20508-20514.
(19) Hileman, R. E.; Fromm, J. R.; Weiler, J. M.; Linhardt, R. J., *Bioessays*, **1998**, *20*, 156-167.
(20) Imberty, A.; Lortat-Jacob, H.; Pérez, S., *Carbohydr. Res.*, **2007**, *342*, 430-439.
(21) Lau, E. K.; Paavola, C. D.; Johnson, Z.; Gaudry, J.-P.; Geretti, E.; Borlat, F.; Kungl, A. J.; Proudfoot, A. E.; Handel, T. M., *J. Biol. Chem.*, **2004**, *279*, 22294-22305.
(22) Shaw, J. P.; Johnson, Z.; Borlat, F.; Zwahlen, C.; Kungl, A.; Roulin, K.; Harrenga, A.; Wells, T. N.; Proudfoot, A. E., *Structure*, **2004**, *12*, 2081-2093.
(23) Murphy, J. W.; Cho, Y.; Sachpatzidis, A.; Fan, C.; Hodsdon, M. E.; Lolis, E., *J. Biol. Chem.*, **2007**, *282*, 10018-10027.
(24) Allen, S. J.; Crown, S. E.; Handel, T. M., *Annu. Rev. Immunol.*, **2007**, *25*, 787-820.
(25) Bernstein, S. L.; Wyttenbach, T.; Baumketner, A.; Shea, J.-E.; Bitan, G.; Teplow, D. B.; Bowers, M. T., *J. Am. Chem. Soc.*, **2005**, *127*, 2075-2084.
(26) Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T., *Nat. Chem.*, **2009**, *1*, 326-331.
(27) Smith, A. M.; Jahn, T. R.; Ashcroft, A. E.; Radford, S. E., *J. Mol. Biol.*, **2006**, *364*, 9-19.
(28) Cole, H. L.; Kalapothakis, J. M. D.; Bennett, G.; Barran, P. E.; MacPhee, C. E., *Angew. Chem.*, **2010**, *122*, 9638-9641.
(29) Severin, I. C.; Gaudry, J. P.; Johnson, Z.; Kungl, A.; Jansma, A.; Gesslbauer, B.; Mulloy, B.; Power, C.; Proudfoot, A. E.; Handel, T., *J. Biol. Chem.*, **2010**, *285*, 17713-17724.
(30) Crown, S. E.; Yu, Y.; Sweeney, M. D.; Leary, J. A.; Handel, T. M., *J. Biol. Chem.*, **2006**, *281*, 25438-25446.
(31) Sweeney, M. D.; Yu, Y.; Leary, J. A., *J. Am. Soc. Mass Spectrom.*, **2006**, *17*, 1114-1119.
(32) Jen, C. H.; Leary, J. A., *Anal. Biochem.*, **2010**, *407*, 134-140.
(33) Schenauer, M. R.; Yu, Y.; Sweeney, M. D.; Leary, J. A., *J. Biol. Chem.*, **2007**, *282*, 25182-25188.

(34) Yu, Y.; Sweeney, M. D.; Saad, O. M.; Crown, S. E.; Handel, T. M.; Leary, J. A., *J. Biol. Chem.*, **2005**, *280*, 32200-32208.
(35) Schenauer, M. R.; Leary, J. A., *Int. J. Mass. Spectrom.*, **2009**, *287*, 70-76.
(36) Seo, Y.; Andaya, A.; Bleiholder, C.; Leary, J. A., *J. Am. Chem. Soc.*, **2013**, *135*, 4325-4332.
(37) Tuinstra, R. L.; Peterson, F. C.; Elgin, E. S.; Pelzek, A. J.; Volkman, B. F., *Biochemistry*, **2007**, *46*, 2564-2573.
(38) Tuinstra, R. L.; Peterson, F. C.; Kutlesa, S.; Elgin, E. S.; Kron, M. A.; Volkman, B. F., *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105*, 5057-5062.
(39) Volkman, B. F.; Liu, T. Y.; Peterson, F. C., *Methods Enzymol.*, **2009**, *461*, 51-70.
(40) Kuloglu, E. S.; McCaslin, D. R.; Kitabwalla, M.; Pauza, C. D.; Markley, J. L.; Volkman, B. F., *Biochemistry*, **2001**, *40*, 12486-12496.
(41) Peterson, F. C.; Elgin, E. S.; Nelson, T. J.; Zhang, F.; Hoeger, T. J.; Linhardt, R. J.; Volkman, B. F., *J. Biol. Chem.*, **2004**, *279*, 12598-12604.
(42) Vogel, G.; Meuleman, D.; Van Dinther, T.; Buijsman, R.; Princen, A.; Smit, M., *J. Thromb. Haemost.*, **2003**, *1*, 1945-1954.
(43) Yu, Y.; Sweeney, M. D.; Saad, O. M.; Leary, J. A., *J. Am. Soc. Mass Spectrom.*, **2006**, *17*, 524-535.
(44) Smith, M.; Thickitt, C., *Rapid Commun. Mass Spectrom.*, **2009**, *23*, 3018-3022.
(45) Cohlberg, J. A.; Li, J.; Uversky, V. N.; Fink, A. L., *Biochemistry*, **2002**, *41*, 1502-1511.
(46) McLaurin, J.; Franklin, T.; Zhang, X.; Deng, J.; Fraser, P. E., *Eur. J. Biochem.*, **1999**, *266*, 1101-1110.
(47) Castillo, G. M.; Lukito, W.; Wight, T. N.; Snow, A. D., *J. Neurochem.*, **1999**, *72*, 1681-1687.

6

Investigating interactions with GAGs: Human β-defensins

Human β-defensins (HBDs) share several structural and functional similarities with chemokines. Previous studies have demonstrated that HBDs can bind glycosaminoglycans (GAGs) and activate some chemokine receptors. This chapter focuses on the binding of human β-defensin-2 (HBD2) and human β-defensin-3 (HBD3) to the model pentameric GAG, fondaparinux (Fx). The work presented here aims to study and characterise the HBD:GAG interactions and aggregation using a combination of biophysical techniques including MS, DT IM-MS, ECD, TEM and isothermal titration calorimetry, to better understand these interactions and determine how they compare to chemokines. In addition the potential of HBD2 and HBD3 to form heterocomplexes is considered.

6.1 Introduction

6.1.1 Chemokines and defensins: similarities in structure and function

The β-defensin subclass of antimicrobial peptides shares a number of structural and functional similarities with the chemokine superfamily of proteins. Defensins are small cationic, cysteine-rich peptides which are divided into three main subfamilies based on the positioning of their disulfide bonds, as discussed in Chapter 1 section 1.4. In β-defensins the connectivity is as follows: Cys^{I}-Cys^{V}, Cys^{II}-Cys^{IV} and Cys^{III}-Cys^{VI}. The positioning of disulfide bonds helps to stabilise the conserved tertiary fold of the β-defensins, which comprises an α-helix and three antiparallel β-strands[1-4]. This tertiary fold is reminiscent of the highly conserved chemokine fold[5,6], however, in chemokines the fold is stabilised most frequently by two disulfide bonds.

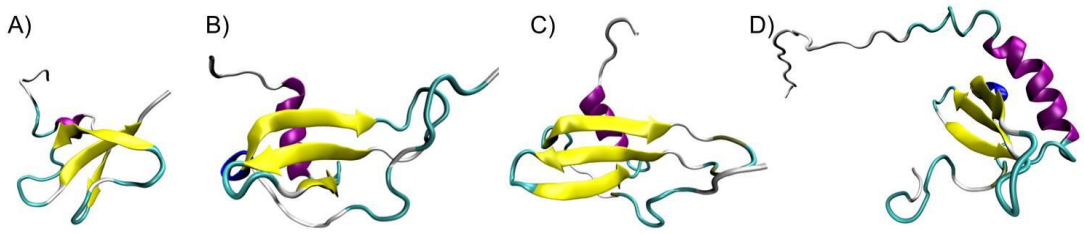

Figure 6.1: Tertiary folds of a HBD and representatives from three major chemokine classes, solved by NMR A) human β-defensin-3 (PDB 1JK6) B) CC chemokine CCL20 (PDB 2JYO) C) CXC chemokine CXCL8, stabilised monomer structure (PDB 1IKL) D) C chemokine lymphotactin, monomeric (PDB 1J9O).

Interestingly, in addition to sharing similar tertiary folds (Figure 6.1), the functions of chemokines and defensins have recently been recognised to overlap. The most well characterised function of defensins is their role within the innate immune system, in which they display antimicrobial properties with broad-spectrum activity towards viruses, fungi, Gram-positive and Gram-negative bacteria[7-9]. In 2000 Krijgsveld *et al*[10] reported the first chemokine displaying antimicrobial activity, following which the majority of human chemokines have since been studied and numerous examples of broad-spectrum antimicrobials have been reported[11-14].

In addition to the recent discovery of antimicrobial properties, chemokines are most well known as signalling proteins in both the innate and acquired immune system[15,16], within which they often interact with G-protein coupled receptors (GPCRs)[5,17]. This is another function which β-defensins share, with human β-defensin-2 (HBD2) and human β-defensin-3 (HBD3) reportedly binding CC-type chemokine receptor-2[18] (CCR2) and CCR6[19] and displaying chemotactic activity[20,21]. Hence it appears as though the functions of defined

chemokine proteins and β-defensin peptides overlap significantly; however, due in-part to the smaller size of β-defensins (2-6 kDa compared with 8-10 kDa for chemokines) and their third disulfide bond, these peptides are not classified as chemokines.

6.1.2 GAG binding

In order to fulfil their full function *in vivo* it is essential for certain chemokines to interact both with their GPCR and with glycosaminoglycans on endothelial cells and in the extracellular matrix[9,22,23], as discussed in Chapter 5. GAG binding is thought to provide a necessary method of localisation and concentration of chemokines and is essential in order to establish a concentration gradient along which cells can migrate during chemotaxis[9].

Numerous antimicrobial peptides have also been shown to interact with GAGs[24], including the defensin family[25-27]. Investigations into defensin:GAG binding are, however, in their infancy in comparison to those carried out on the chemokine superfamily. To-date most studies considering GAG binding have focused on the implications with respect to antimicrobial activity, with few considering chemotaxis[28], despite overlapping properties of these two families. One such study demonstrated that a member of the α-subclass of defensins can bind the GAG dermatan sulfate, which subsequently reduced the antibacterial effect of the defensin[25]. McCullough *et al*[26] studied interactions between a heparin-derived sulfated disaccharide and a series of β-defensin related peptides using mass spectrometry (MS) and ion mobility-mass spectrometry (IM-MS). A number of monomeric defensin plus GAG complexes were observed here which were found to adopt compact collision cross sections (CCS), suggesting tight specific binding. Furthermore, as with the α-defensins, GAG binding decreased the antimicrobial activity of these β-defensin related peptides.

In 2010 Seo *et al*[29] presented the first detailed structural study into the GAG binding ability of human β-defensins. Here investigation focused on interactions between HBD2 and a series of GAGs using a combination of biophysical techniques, including gel mobility shift assays, mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy. The results highlighted that HBD2 was capable of binding GAGs of a chain length greater than two; however, above six units extensive aggregation was observed. Both MS and NMR studies highlighted that HBD2 forms a dimer upon binding with the pentameric GAG fondaparinux (Fx), consistent with observations made for chemokines which have been reported to bind GAGs most frequently in monomeric or dimeric forms[30-33] in *in vitro* studies.

Interactions between GAGs and the remaining human-β-defensins have been under-reported. The investigations presented here aim to address this through studying the interactions of both HBD2 and HBD3 with the model GAG Fx, using a combination of biophysical techniques. The interactions between HBD2 and Fx are further characterised using drift tube ion mobility-mass spectrometry (DT IM-MS), electron capture dissociation (ECD) and isothermal titration calorimetry (ITC). The ability of HBD3 to bind Fx and stoichiometry of the interactions is discerned from MS. The conformations of the resulting HBD3+Fx complexes are explored using DT IM-MS. The extent of GAG-induced aggregation of HBDs has not to-date been investigated and hence the aggregation occurring upon Fx addition to both of these β-defensins is studied using transmission electron microscopy (TEM). The combination of techniques presented here aims to shed light on this intricate biological process and further understand these interactions, enabling any similarities or differences in comparison to chemokine:GAG binding to be discerned.

6.1.3 Heterodimerisation of chemokines and implications for HBDs

Many chemokines form dimers *via* distinct, defined dimerization motifs generally represented by either the CC or CXC chemokine dimer interface; however, novel dimers are also known[34,35]. In general chemokines of the CC subclass dimerise through formation of a two-stranded antiparallel β-sheet near the N-termini, whilst chemokines of the CXC subclass dimerise through the first strand of the monomeric β-sheet[36]. The characterisation of defined dimerization motifs for the CC and CXC classes of chemokines suggested that chemokines of the same subclass could heterodimerise using these motifs and indeed heterochemokine dimers have been observed[37], in some cases having higher affinity over homodimers[38,39]. Furthermore, it has been suggested that heterodimeric chemokine complexes could have functional consequences, adding another layer of complexity to these intricate biological pathways and potentially providing an additional mechanism for *in vivo* modulation[38,40-42].

Studies have shown a correlation between GAG binding and chemokine homodimerisation and oligomerisation[33,43]; hence, recent studies have considered the possibility of GAG modulation of heterodimerisation[44]. Crown *et al*[44] investigated a number of CC chemokines and their propensity for heterodimerisation by MS and NMR, reporting several CC heterodimers, which in the case of CCL2 and CCL8 reduced the homodimer intensity, highlighting that these heterodimeric complexes can be highly favoured. By NMR it was found that CCL2+CCL8 dimerised through formation of a CC dimer interface, signifying the specificity of this heterodimeric chemokine. Significantly, heterochemokine complexes were observed to bind Fx. Furthermore, CCL8 and CCL11, which did not interact in the absence

of Fx were observed to form both CCL8+CCL11 and CCL8+CCL11+Fx complexes in the presence of Fx, implicating GAGs in the regulation of chemokine heterodimerisation and suggesting heterodimers could be of functional relevance.

Both HBD2[1] and HBD3[3] have been found to form homodimers, although the dimeric structure of HBD3 has not to-date been solved. The ability of different members of the β-defensin subfamily to form heterodimeric complexes has not yet been reported. The propensity of HBD2 and HBD3 to form homodimers, coupled with the fact that they interact with the same CCRs[18] and can be found in the same regions[7,8,45-47] makes them attractive targets to investigate the possibility of heterodefensin complex formation. In order to consider heterodefensin complex formation MS and DT IM-MS studies are performed in the absence and presence of Fx.

6.2 Experimental

6.2.1 Peptide and glycosaminoglycan samples

HBD2 and HBD3 produced by solid phase synthesis were purchased from the Peptide Institute (Osaka, Japan). Stock solutions of 200 μM peptide in 20 mM AmAc were prepared and aliquots were stored at -20 °C until use. Ammonium acetate salt was purchased from Fischer Scientific (Loughborough, UK) and prepared in high purity water, in the resistivity range of 18.0-18.2 MΩ•cm obtained from an Arium 611 water purification system (Satorius, Göttingen, Germany).

Bovine pancreatic trypsin inhibitor (BPTI) was purchased from Sigma-Aldrich Company Ltd (Dorset, UK). A stock solution at 1 mM protein concentration was prepared in deionised water and stored in aliquots at -20 °C until use.

Concentrations were verified using the Pierce BCA concentration assay (Thermo Fischer Scientific, Wilmington, DE, USA) following manufacturer's instructions but using volumes 10 times lower than stated in the manual, afforded through the use of low volume UV cuvettes.

Foundaparinux sodium was a gift from GlaxoSmithKline (Brentford, UK). The foundaparinux solution was prepared at a stock concentration of 587 μM in water. Aliquots of which were dialysed overnight versus water using a 1,000 molecular weight cut-off Micro DispoDialyzer (Harvard Apparatus, MA, USA), and stored at -20 °C until use.

6.2.2 Sample preparation for MS studies

In all cases samples were prepared in 20 mM AmAc at pH 6.8.

Individual HBD2 and HBD3 samples were prepared at 50 μM peptide concentration. HBD2 plus Fx and HBD3 plus Fx solutions were prepared at 1:1 molar ratios at 50 μM peptide concentration. HBD3 plus Fx was additionally studied at a 2:1 defensin:GAG ratio at 50 μM HBD3 concentration.

For HBD2 plus HBD3 MS binding studies 1:1 mixtures were prepared at 50 μM total peptide concentration and for MS binding studies of HBD2 plus HBD3 plus Fx, mixtures were prepared at a 1:1:1 ratio with each component being present at 25 μM.

For MS studies BPTI was prepared at 50 μM. BPTI plus Fx samples were prepared at a 1:1 ratio at 50 μM BPTI concentration. HBD2 plus BPTI was prepared at a 1:1 ratio at 50 μM total protein concentration. HBD2 plus BPTI plus Fx was prepared at a 1:1:2 ratio at a 50 μM total protein concentration plus 50 μM Fx.

All MS studies were performed on a Q-TOF Ultima (Waters, Manchester, UK), for operating conditions see Chapter 2 section 2.2. Collision induced dissociation experiments, HBD2 plus HBD3 and HBD2 plus BPTI studies were performed post-high mass upgrade and all other experiments were performed prior to the high mass upgrade.

6.2.3 Sample preparation for DT IM-MS studies

For determination of CCS for all HBD2 and HBD2+Fx species, 50 μM sample solutions of HBD2 plus Fx at a 1:1 ratio was prepared. In addition a 'supercharged' sample was also studied prepared at the same concentration, however, in this case the buffer was 20 mM AmAc plus 1 % *m*-NBA. For determination of CCS of HBD3 and HBD3+Fx species to limit aggregation sample solutions were prepared at a 2:1 HBD3:Fx ratio, with HBD3 at 50 μM concentration. Additionally, one repeat of HBD2 and HBD3 prepared as individual peptide solutions at concentrations of 50 μM were acquired as a control.

For HBD2+HBD3 complexes CCS were determined using 1:1 mixtures at 100 μM total peptide concentration. For HBD2+HBD3+Fx complex CCS determination, mixtures were prepared at a 2:2:1 ratio (HBD2:HBD3:Fx) at a total peptide concentration of 100 μM.

All DT IM-MS measurements were performed on an in-house modified Q-TOF (Micromass UK Ltd.), described in Chapter 2 Section 2.3.

6.2.4 Sample preparation for ECD studies

For HBD2+Fx ECD studies, samples were prepared at a 1:1 ratio at 30 μM peptide concentration in 20 mM AmAc. In addition a 'supercharged' sample was also studied prepared at the same concentrations, however, in this case the buffer was 20 mM AmAc plus 1 % *m*-NBA. Finally, a HBD2 sample was prepared at 25 μM concentration in 20 mM AmAc.

All ECD studies were performed on a 12T Apex Ultra Qh FT-ICR mass spectrometer (Bruker Daltonik Gmbh, Bremen, Germany), details on the instrument and its operation can be found in Chapter 2 section 2.5.

6.2.5 Sample preparation for isothermal titration calorimetry studies

For isothermal titration calorimetry **(**ITC) experiments on HBD2 plus Fx, HBD2 was prepared either at 10 or 15 μM (cell solution) and Fx solutions prepared at 175 or 88 μM (syringe solution) in 20 mM AmAc. To ensure buffer composition was identical for each sample and to limit any buffer dilution effects in the ITC experiments Fx was dialysed for four hours against 12 L of 20 mM AmAc, an aliquot of dialysis buffer was then used to prepare the HBD2 sample. Control experiments were carried out in which Fx (88 μM) was injected into a cell filled with buffer and buffer was injected into a cell containing HBD2 (10 μM).

All ITC experiments were performed using a MicroCal Auto-ITC_{200} (GE Healthcare and Life Sciences, Little Chalfont, UK), details on the instrument and its operation can be found in Chapter 2 section 2.8.

6.2.6 Sample preparation for TEM studies

For TEM studies HBD2 plus Fx and HBD3 plus Fx sample solutions were prepared at 1:1 ratios at a peptide concentration of 50 μM. HBD3 plus Fx was additionally studied at a 2:1 (HBD3:GAG) ratio, at 50 μM peptide concentration. The HBD2 plus HBD3 plus Fx sample was prepared at a 1:1:1 ratio with each component present at 50 μM concentration. Additionally, HBD3 prepared at 65 μM was studied as a control. All TEM samples were prepared in 20 mM AmAc. For details on TEM grid preparation procedure see Chapter 2 section 2.7.

All TEM experiments were performed on a Philips CM120 Biotin transmission electron microscope (Philips, Eindhoven, Netherlands), details on the instrument and its operation can be found in Chapter 2 section 2.7.

6.2.7 Molecular dynamics simulations

All molecular dynamics simulations and theoretical calculations discussed in this chapter were performed by Dr Massimilano Porrini. Details on the procedure can be found in brief in Appendix 5 section A5.1. Theoretical CCS were determined for *in vacuo* minimised monomer and dimer+Fx of HBD2. The structure of Fx was taken from the PDB file of its complex with Antithrombin-S195A factor Xa (PDB 2GD4). Initial coordinates for HBD2 were taken from the crystal structure of the dimer (PDB 1FD3).

6.3 Results and discussion

6.3.1 Ability of HBD2 to bind Fx

6.3.1.1 Stoichiometry of binding: MS studies

Previous MS studies from the Barran group highlighted that HBD2 can bind Fx, with dimeric HBD2 binding a single Fx chain[28,29]. These studies were, however, performed at relatively high concentration with samples prepared at a 1:1 ratio at 135 µM HBD2 concentration in 20 mM ammonium acetate (AmAc). The studies presented in this chapter aim to further probe and characterise these interactions and the complexes formed. Initial studies into HBD2 binding reported here focused on reproducing the MS binding studies on different instrumentation using a lower concentration of both HBD2 and Fx, samples were therefore prepared at a 1:1 ratio at 50 µM peptide concentration in 20 mM AmAc. In addition, n-ESI capillaries, source voltages and pressure were tuned to enable maximum signal transmission, enabling the spectra shown in Figure 6.2 to be acquired.

The spectrum acquired for HBD2 (Figure 6.2A) establishes that HBD2 presents mostly in a monomeric form, consistent with the findings from solution-phase NMR studies[3], and is capable of accepting between three and five protons. In addition some lower intensity dimeric (charge states 4+ and 5+) and trimeric (charge states 6-8+) species are also detected, attributed to HBD2 oligomerising in solution at this concentration and all observed masses are in good agreement with the expected masses (Appendix 5 Table A5.1). Dimeric HBD2 has been previously reported in X-ray crystallography studies[1], although at much higher

concentrations than those used here; furthermore, a trimeric HBD2 structure has not been reported. The results suggest that either conditions used here enable low intensity species to be preserved, transmitted and detected, or alternatively conditions here are promoting their formation.

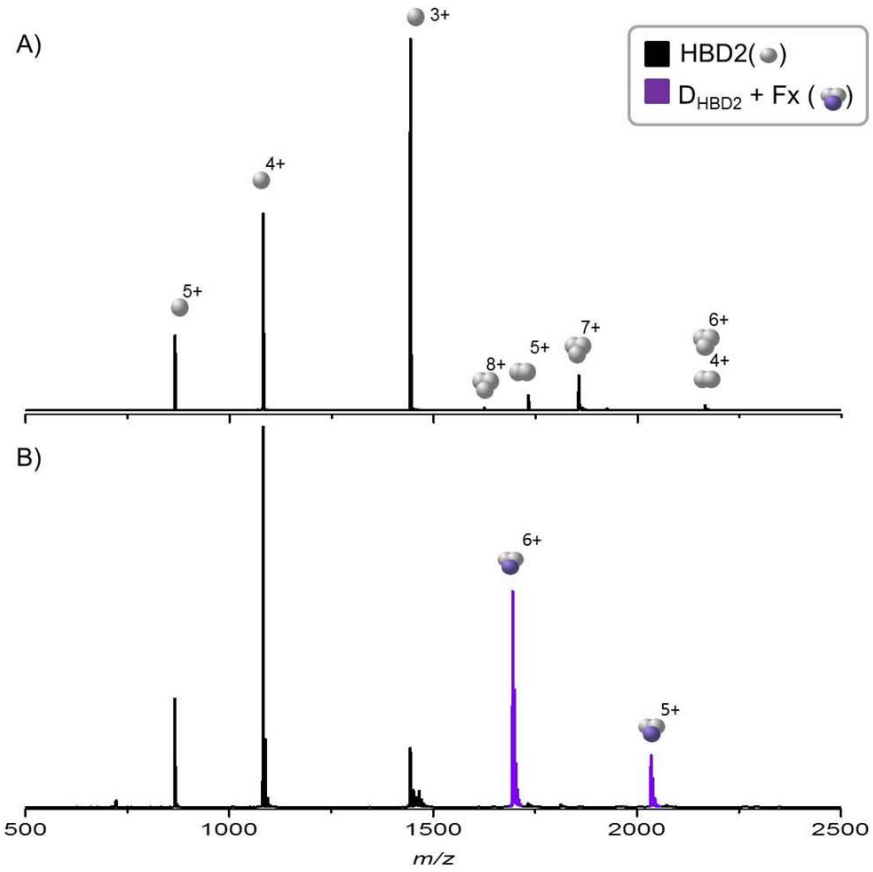

Figure 6.2: Mass spectra of HBD2 A) 50 µM HBD2 B) HBD2 plus Fx prepared at a 1:1 ratio at 50 µM peptide concentration.

Upon addition of stoichiometric concentrations of Fx, two high intensity GAG-bound complexes are observed (Figure 6.2B), both of which are consistent with dimeric HBD2 binding a single chain of Fx (D_{HBD2}+Fx), presenting in charge states 5+ and 6+. The observed masses are in good agreement with the expected masses (Appendix 5 Table 5.1). It is interesting to note that in its Fx-bound form, HBD2 exclusively forms dimers with no other Fx-bound species being observed. It has been suggested previously for chemokines that the GAG chain length is an important factor in the consideration of complexes formed, with shorter GAGs often giving rise to monomeric complexes which have been attributed to the shorter length being insufficient to bridge the interface[30,31]. It appears the pentameric GAG Fx is sufficient to bridge the dimeric interface of HBD2, consistent with previous studies[28,29]. The high intensity of these species highlights the stability and specificity of the complexes,

proposing that Fx preferentially stabilises the dimeric interface between these two highly cationic peptides. It is important to note that upon addition of Fx there was evidence of aggregation with visible aggregates observed in the n-ESI capillaries and blocking with time, similar to observations made for the chemokine Ltn upon addition of Fx (Chapter 5).

Interestingly, upon addition of Fx, the distribution of unbound HBD2 is altered with respect to oligomer population, with no unbound dimers or trimers present, signifying Fx is shifting this equilibrium. Alternatively, their absence in the spectrum could be due to the lower signal intensity here, due to the aggregating HBD2 plus Fx sample, which may cause these lower intensity peaks to no longer be observed. The charge state distribution of monomer is also affected upon addition of Fx, the population of $[M_{HBD2}+3H]^{3+}$ is considerably diminished and $[M_{HBD2}+4H]^{4+}$ is observed to be the most intense. This could be attributed to a number of factors, such as the addition of Fx may be perturbing the conformations and hence charge state distribution of unbound monomeric species present, either through promotion of more extended species or alteration of the solution pH. In order to discount this, solution pH was measured after addition of Fx and remains constant, enabling pH effects to be disregarded. The promotion of more extended conformations in the presence of Fx can be studied *via* DT IM-MS and is considered in detail below.

An alternative explanation is that Fx is binding to the species in solution which are capable of accepting three protons and hence diminishing their relative intensity; considering the charge states of the bound dimers (5+ and 6+) this could indeed hold true. To probe this possible explanation, and the stability of the D_{HBD2}+Fx species, CID studies were performed on the 5+ and 6+ species. CID experiments showed that the first fragmentation pathway is the loss of SO_3 groups from Fx, highlighting the stability of this complex (Appendix 5 Figure A5.1). The complex can, however, be dissociated through increasing the collisional energy and $[M_{HBD2}+3H]^{3+}$ species are then observed supporting the hypothesis that these species are involved in Fx binding and providing a plausible explanation for the lower observed relative intensity of these species in the presence of Fx.

6.3.1.2 Conformations of HBD2:Fx complexes: insights from DT IM-MS

Conformations adopted by HBD2 and D_{HBD2}+Fx were then studied using DT IM-MS. In both cases CCS were determined by studying samples prepared at a 1:1 ratio of HBD2:Fx. Considering first the unbound species, CCS of monomeric HBD2 were determined over the charge state range $[M_{HBD2}+3H]^{3+}$ to $[M_{HBD2}+6H]^{6+}$ (Figure 6.3 and Appendix 5 Table A5.2).

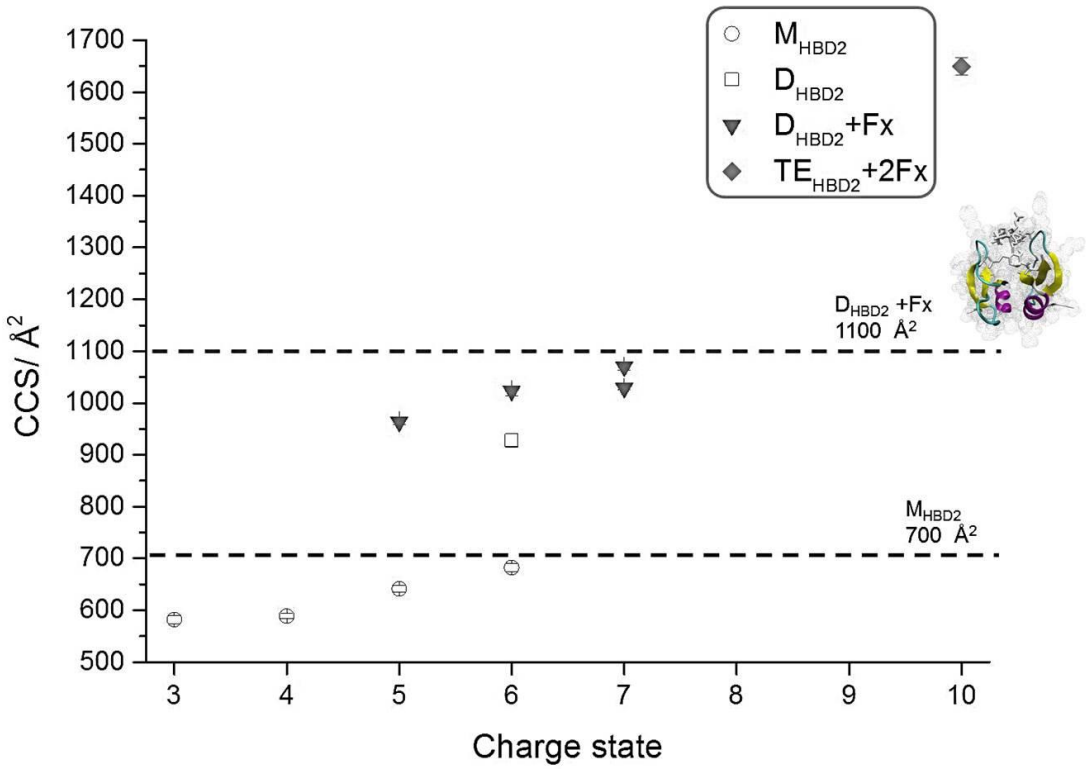

Figure 6.3: Average CCS of HBD2, and Fx bound HBD2, calculated from three repeats, error bars represent the standard deviation between these values. Unbound species (hollow symbols) were determined in the presence of Fx. Fx bound species of dimeric (D) and tetrameric (TE) HBD2 are also shown here (filled symbols). Theoretical CCS for monomeric HBD2 and dimeric HBD2+Fx (structure shown in insert) are represented by black and grey dashed lines respectively.

CCS of the monomeric species are very similar, presenting over a narrow conformational range increasing from ~582-683 Å^2, consistent with a single conformational family, with the slight increase observed being attributed to Coulombic repulsion of proximal charges. The presence of a single, stable conformational family is accredited to the three disulfide bonds present within HBD2 which would act as conformational staples, restricting the fold[48]. The theoretical CCS was determined for monomeric HBD2 by clipping a single monomer from the HBD2 dimer, solved by X-ray crystallography, followed by gas-phase minimisation. The resulting theoretical CCS (700 Å^2) is in excellent agreement with the experimental CCS of monomeric HBD2, suggesting that the peptide ions could be adopting conformations similar to this.

In order to determine if the conformations and hence CCS of monomeric HBD2 are altered in the presence of Fx, CCS were determined for a single repeat of HBD2 in 20 mM AmAc and were found to lie within experimental error of those determined in the presence of Fx (Appendix 5 Table A5.2) and in good agreement with previously reported values[28]. This signifies that Fx does not alter the conformation of unbound HBD2 and hence the shift in the MS charge state distribution for unbound HBD2 in the presence of Fx is most likely due to

Fx binding to the species in solution which are capable of accepting three net protons and hence reducing their relative intensity consistent with findings from CID (Appendix 5 Figure A5.1).

CCS of the GAG-bound complexes (D_{HBD2}+Fx) were then determined over the charge state range 5+ to 7+ (Figure 6.3 and Appendix 5 Table A5.3). D_{HBD2}+Fx exists in a small range of CCS (~964-1070 $Å^2$), suggesting it is in a similar, stable conformation at all charge states, which is not greatly affected by increasing charge. The CCS for a single charge state (6+) of unbound dimeric HBD2 could be determined here and is smaller by ~10 % than the Fx-bound dimers. The similarities in CCS between D_{HBD2} and D_{HBD2}+Fx indicates specific GAG-bound complexes are formed, as non-specific complexes would be expected to present with much larger CCS, consistent with the findings from CID which suggested stable, specific complexes were formed.

The structure of the D_{HBD2}+Fx complex has been studied by NMR, identifying a number of residues thought to be involved in binding[29]; however, a solved complex structure has not yet been reported. In order to further probe the conformation of this complex, detailed molecular dynamics (MD) studies were performed in combination with the work presented in this thesis, with the aim of predicting the conformation of the bound complex, *via* theoretical CCS comparison. In this study the initial coordinates for dimeric HBD2 were taken from the structure solved by X-ray crystallography (PDB 1FD3) and coordinates of Fx were taken from a complex with antithrombin and coagulation factor Xa, (PDB 2GD4). The structure of the D_{HBD2}+Fx complex was predicted using Autodock 4.2 software[49]. The docking simulations provided a likely candidate complex. The complex was then minimised *in vacuo* and subjected to gas-phase dynamics prior to CCS determination. The resulting candidate complex is shown in Figure 6.4, in which the Fx chain binds between two HBD2 monomeric units with multiple basic sites of interaction, which would strengthen and stabilise this complex. Significantly, the residues which are shown to be involved in binding, compare remarkably well to the residues implicated in binding through NMR studies into HBD2 plus Fx[29] (Figure 6.4B). The CCS of this species (1100 $Å^2$) is in excellent agreement with the experimental CCS obtained, suggesting the complex formed could indeed be adopting this conformation and representing the first predicted structure for a defensin:GAG complex.

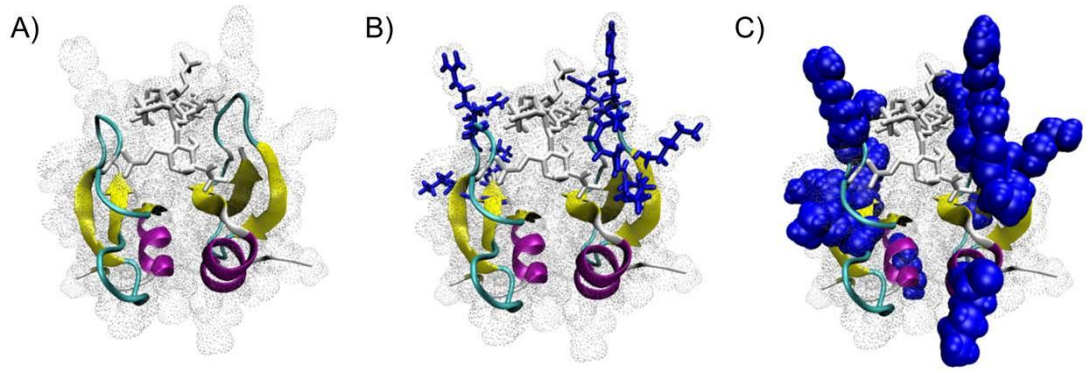

Figure 6.4: Theoretical D_{HBD2}+Fx complex obtained from MD simulations, initial coordinates of dimeric HBD2 were obtained from PDB 1FD3. A) No residues labelled B) All basic residues identified in the previous study by Seo *et al*[29] to interact with Fx are shown in blue. C) All basic residues in in HBD2 shown as a surface. All structures for this figure were visualised in VMD.

In addition, there is evidence from DT IM-MS that HBD2 can form tetrameric complexes binding two Fx chains and presenting in a 10+ charge state (Figure 6.3). This species is *m/z* coincident with $[D_{HBD2}+Fx+5H]^{5+}$, however, they can be separated out based on their mobility through the buffer gas filled drift cell. The CCS was determined as 1649 ± 17 Å^2, just under twice the size of the D_{HBD2}+Fx species suggesting it could be a dimer of dimers.

6.3.1.3 HBD2:GAG binding interface: ECD studies

Considering D_{HBD2}+Fx in more detail, the previously reported NMR study into interactions between HBD2 and Fx[29] distinguished a number of chemical shift perturbations in HBD2 including; R22, R23, K25, K39 and K40, all of which fall within the binding interface of the gas-phase complex predicted through MD simulations (Figure 6.4B). Interestingly, GAG binding in chemokines can involve a binding motif of sequence BBXB (where B is a basic amino acid)[50] which is consistent with the ^{22}RRYK^{25} sequence implicated in HBD2:GAG binding by NMR[29]. Oligomerisation of chemokines and defensins on GAG binding, however, can complicate the identification of such regions and hence consideration of purely linear sequences alone is not always applicable. In order to probe this interface further in the gas-phase, ECD studies were performed with the aim that the high sensitivity of the FT ICR-MS instrument used here would enable even low intensity fragmentation to be detected.

Samples were initially prepared in 20 mM AmAc, enabling D_{HBD2}+Fx at charge states 5+ and 6+ to be transferred and detected at high enough intensity to be subsequently isolated and subjected to ECD. For $[D_{HBD2}+Fx+5H]^{5+}$ no fragmentation is observed, attributed to its low charge state and compact nature (Appendix 5 Figure A5.2A). $[D_{HBD2}+Fx+6H]^{6+}$ also remains mostly intact; however, here limited backbone cleavage occurs (Appendix 5 Figure A5.2B), with fragments arising from cleavage at either termini whilst the interface remains

intact. The lack of fragmentation observed at both charge states studied is consistent with a compact species (as shown by DT IM-MS), stabilised through numerous interactions. Furthermore, the three disulfide bridges in HBD2 would be preferentially cleaved in ECD prior to backbone fragmentation, limiting the fragmentation efficiency. The charge state of the species also has to be considered as ECD is known to be more effective at higher charge states[51]. It is therefore conceivable that here the low charge states of the complexes are limiting the fragmentation efficiency.

In an attempt to further probe the stability of this complex and its binding interface a 'supercharging' reagent known as *meta*-nitrobenzyl alcohol (*m*-NBA)[52] was added to the sample solution (1 % in 20 mM AmAc). Addition of *m*-NBA promotes higher charge states of monomeric HBD2 and D_{HBD2}+Fx (Appendix 5 Figure A5.3A). Previous studies have shown that addition of *m*-NBA can disrupt native protein complexes[53] and perturb the conformations of monomeric systems (see Chapter 4 section 4.3.3). Therefore DT IM-MS experiments were first performed on the supercharged samples to discern if addition of this reagent alters the conformations of the monomeric peptide or the complex. CCS determined from the supercharged sample (Appendix 5 Table A5.4) are within experimental error of those determined in the buffered sample (Appendix 5 Tables A5.2 and A5.3), highlighting that here *m*-NBA is not altering the conformation of the peptide or the complex. Addition of *m*-NBA enabled the D_{HBD2}+Fx complex to be detected in charges state 6-9+ on the DT IM-MS instrument, whilst in buffered conditions only charge states 5-7+ were observed. The CCS for the 8+ and 9+ species are slightly more extended (1175 ± 24 $Å^2$ and 1256 ± 15 $Å^2$ respectively) than those of the lower charge states, as expected. These more extended and unfolded species therefore present favourable targets for further ECD analysis. Supercharged species were then studied on the FT ICR-MS instrument, enabling two higher charge states of the complex to be transmitted at high enough intensity to be exposed to ECD fragmentation, namely $[D_{HBD2}+Fx+7H]^{7+}$ and $[D_{HBD2}+Fx+8H]^{8+}$ (Appendix 5 Figure A5.3B).

$[D_{HBD2}+Fx+7H]^{7+}$ undergoes low intensity ECD fragmentation (Appendix 5 Figure A5.4), producing a number of fragment ions, corresponding to the D_{HBD2}+Fx complex fragmenting at the first and second amino acids of the N-termini and the last amino acid at the C-termini of a HBD2 monomer, with the complex remaining intact. In addition, at this charge state there is evidence for complex dissociation to monomer. For $[D_{HBD2}+Fx+8H]^{8+}$ the main fragment species observed are due to the complex remaining intact and fragmenting at the termini, with the most extensive fragmentation being in the first six amino acids of the N-termini. Interestingly at this charge state no dissociation of the complex is observed,

suggesting this complex is more stable. The buffered DT IM-MS experiments showed $[D_{HBD2}+Fx+7H]^{7+}$ adopts two conformations (Figure 6.3); a compact and more extended form suggesting it is a key transitional charge state between conformational families.

This limited complex dissociation and observation that covalent bonds in the termini are the favoured cleavage site demonstrates the stability of this complex. The findings for this complex are consistent with the complex structure shown in Figure 6.4 which predicts multiple interactions with Fx along the backbone of both HBD2 monomers in the dimer which would produce a stable, fragmentation-resistant species. The lack of fragmentation along the peptide backbone in all complex species is attributed in part to the stability of HBD2, with each monomeric unit in the dimer containing three constraining disulfide bonds which stabilise the tertiary fold and hence limit fragmentation. Furthermore, in ECD disulfide bonds are preferentially reduced through the addition of the electrons, which significantly limits any backbone fragmentation for disulfide bridged proteins. ECD studies were also performed on HBD2 monomer and very little fragmentation is observed, with the majority of fragmentation occurring in the first 13 amino acids of the N-termini consistent with a stable, constrained species (Appendix 5 Figure A5.5). Similar comparisons with dimeric HBD2 could not be made due to its low abundance. The ECD studies, in combination with DT IM-MS and CID, suggest that HBD2 forms stable, compact specific interactions with Fx to form a dimeric HBD2 plus Fx complex.

6.3.1.4 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were then performed with the aim of confirming the stoichiometry of this interaction in solution and to determine the dissociation constant (K_d) of binding. In all ITC experiments 16 injections were performed, with the first being a lower volume, as the error in the first injection is often the highest and therefore this injection is routinely discounted for analysis.

ITC experiments involved injecting Fx into a cell filled with HBD2 and measuring the resulting change in heat. Two control experiments were also run; injecting buffer into a cell of HBD2 and injecting Fx into a cell filled with buffer. These control experiments enable any heat change due to buffer mismatch effects to be identified and accounted for, here each control experiment shows no such effects enabling them to be discounted (Figure 6.5). The ITC binding study was performed by injecting 88 µM Fx into a cell filled with 10 µM HBD2 (Figure 6.5). It is clear that there are at least two different physical processes occurring here. At low molar ratios the process is dominated by endothermic events; however, as the molar

ratio is increased, the overall reaction is dominated by exothermic events, characterised by the switch in peak direction.

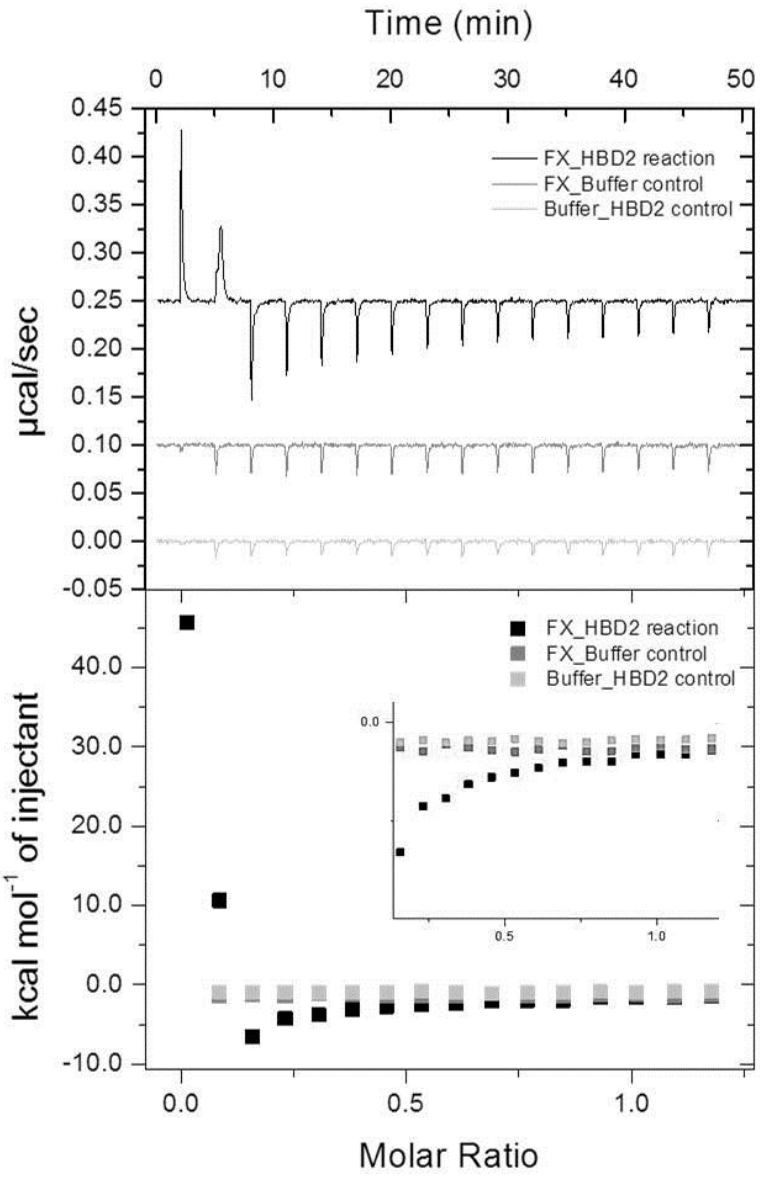

Figure 6.5: ITC traces obtained for HBD2 plus Fx reaction, obtained at 15 µM HBD2 concentration and 88 µM Fx concentration. Fx plus buffer control (injecting 88 µM Fx into a cell filled with buffer) and buffer plus HBD2 control (injecting buffer into a cell filled with 10 µM HBD2). Data processed to correct baseline, normalise and subtract controls. Insert shows only exothermic processes. In all cases buffer was 20 mM AmAc.

A similar effect was detected in experiments in which 175 µM Fx solution was injected into a cell filled with 10 µM HBD2, highlighting this observation is repeatable. One possible explanation is that at low molar ratios the process is dominated by electrostatic interactions attracting a defensin, or multiple defensins, towards a sugar molecule and disrupting the water core around the sugar and/or defensin, and displacement of counter-ions[54] bound to Fx and remaining after dialysis, both of which would contribute towards an endothermic

signature. When the molar ratio is increased these forces no longer dominate and the molecules bind in greater numbers forming the complex, creating the exothermic signature. It is, however, likely that there are many processes occurring at once, complicating this signature further and therefore at this point it is not possible to conclusively determine the processes occurring. For example the aggregation occurring upon addition of Fx would also complicate these results.

No standard ITC model can be fitted to this data and hence the K_d and stoichiometry of the reaction, as well as determination of the multiple processes occurring here cannot be definitively determined. In order to overcome this, highly robust data would be required involving multiple repeats, and more injections per repeat. Due to the sample-intensive nature of ITC, with each binding study requiring 60 µL of Fx and 350 µL of HBD2 per run and the same volume being required for controls which are necessary for each individual run, further ITC studies are not feasible here.

Clearly the binding process between HBD2 and Fx is complex and more experiments are needed here to fully characterise these interactions, the ITC observations are repeatable but are dissimilar to those made for previous chemokine:GAG binding events which showed purely exothermic processes[31]. Despite this, information can still be gleaned from these experiments; HBD2 and Fx do bind in solution and an end point can be observed, based on the plateau of peaks. Considering only the exothermic processes, and hence the later three quarters of the data, fitting to a standard ITC model a binding stoichiometry of ~0.5 is determined which is consistent with a dimer of HBD2 binding a single Fx chain, in agreement with findings from MS. Fitting only three quarters of the data is evidently not robust and therefore additional studies in this area are required, which as discussed earlier necessitates substantial material and therefore alternative sources and methods of production (such as peptide expression) would be vital to generate the necessary material.

6.3.1.5 Assessing aggregation: TEM studies

MS studies into the interactions of HBD2 with Fx highlighted that aggregation can occur upon addition of this GAG, as has been observed with chemokines[23,43] including Ltn (Chapter 5). In order to further study this phenomena transmission electron microscopy (TEM) studies were performed. A 1:1 mixture with Fx (Figure 6.6), yields large globular aggregates, dissimilar in appearance to the ribbon-like species observed for WT and CC3 Ltn (Chapter 5 section 5.3.1 and 5.3.3) but similar in morphology to those observed previously for WT 1-72 Ltn (Chapter 5 section 5.3.2). The aggregates observed for HBD2 are,

however, smaller and generally fewer in number then the WT and WT 1-72 Ltn species, either due to HBD2 having a lower affinity for the grid or aggregation occurring to a lesser extent, which is consistent with results from MS. The observation of these species is significant, however, clearly emphasising that aggregation is occurring here.

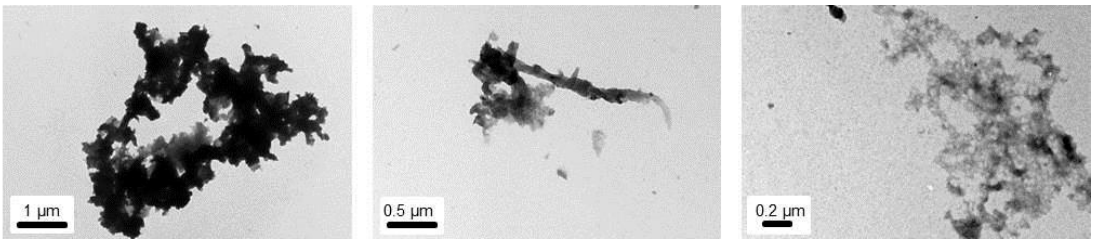

Figure 6.6: Representative TEM images obtained for 1:1 HBD2 plus Fx (50 µM).

6.3.2 Ability of HBD3 to bind Fx

The results from the previous sections demonstrate that HBD2 can bind Fx, producing stable specific dimeric complexes and displaying similar traits to the chemokine superfamily of proteins. In order to further probe the GAG binding affinity of β-defensins and their similarities to chemokines another member of this subfamily of peptides was studied, human β-defensin-3 (HBD3).

6.3.2.1 Stoichiometry of binding: MS studies

MS studies were performed to determine if HBD3 could bind Fx and to discern the stoichiometry of any resulting complexes. The spectrum obtained for HBD3 peptide solution showed HBD3 exists mostly in a monomeric form (M_{HBD3}) under the conditions employed here (Figure 6.7A), in the charge state range $[M_{HBD3}+4H]^{4+}$ to $[M_{HBD3}+7H]^{7+}$ with the species accepting five charges being the most intense. In solution HBD3 has been shown to exist in a dimeric form[3]; however, these studies were performed at 360 µM, and therefore the high intensity of monomer observed here is attributed to the substantially lower concentration used (50 µM). A low intensity peak was assigned as dimeric HBD3 (D_{HBD3}), in the 7+ charge state, demonstrating that dimeric species can form. The expected and observed masses for all species are in good agreement (Appendix 5 Table A5.5).

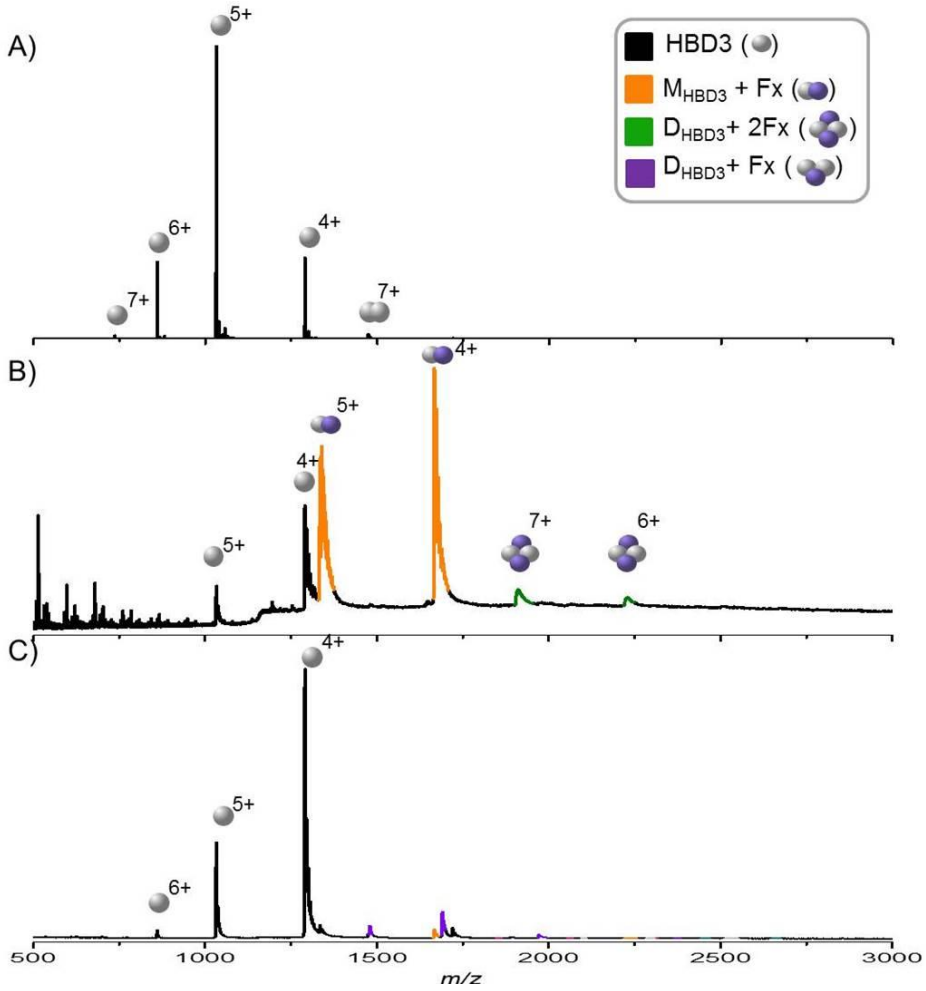

Figure 6.7: Mass spectrum of HBD3 A) 50 µM HBD3 B) HBD3 plus Fx prepared at a 1:1 ratio at 50 µM peptide concentration. C) HBD3 plus Fx prepared at a 2:1 ratio at 50 µM HBD3 concentration.

Previous MS-based studies showed that HBD3 aggregates extensively upon the addition of equimolar concentrations of Fx, resulting in an inability to record usable spectra and observe binding[28]. In order to overcome this, a number of conditions were tuned to optimise the signal of oligomers, including pressure in the source region and tuning of the n-ESI capillaries pulled for use in these studies. Through optimisation of these parameters it was possible to obtain a mass spectrum for HBD3 plus Fx at a 1:1 ratio at 50 µM concentration, as shown in Figure 6.7B. The spectrum shows signs of extensive aggregation as exemplified by the raised baseline particularly over the *m/z* range 1100-3000. Despite the substantial aggregation occurring, a number of peaks are distinguished due to HBD3 binding to Fx, most notably monomeric HBD3 is observed at high intensity to bind a single Fx chain. The complex presented in both 4+ and 5+ charge states and is observed to be the most intense species in the spectrum. In addition, two lower intensity peaks could be assigned to dimeric HBD3 binding two chains of Fx (D_{HBD3}+2Fx), in 6+ and 7+ charge states. $[D_{HBD3}+2Fx+6H]^{6+}$ is mass coincident with $[M_{HBD3}+Fx+3H]^{3+}$ and this peak is most likely

due to both species. Previous studies into the GAG binding propensity of HBD3 showed tetrameric (heparin degree of polymerisation (dp) 4) and hexameric (dermatan sulfate dp6) GAG chains can bind to both monomeric and dimeric HBD3[28], consistent with the ability of Fx to bind in both monomeric and dimeric complexes as observed here. The presence of both monomeric and dimeric bound species was attributed to the distance between the GAG binding site on each HBD3 monomer being greater than on HBD2[28], which only formed dimeric GAG bound complexes with the same GAGs.

In addition to the Fx-bound species, two further low intensity peaks are observed due to unbound monomeric HBD3. The peaks present at the low *m/z* region are due to salts and contaminants from the Fx sample. Despite the low quality of the spectrum obtained here, due to extensive aggregation, it is clear that HBD3 can bind Fx. In order to obtain a spectrum of higher quality, the concentration of Fx was lowered and a sample was prepared at a 2:1 ratio of HBD3:Fx at 50 µM peptide concentration (Figure 6.7C and Figure 6.8). Here aggregation is limited, enabling a clearer spectrum to be obtained. The most intense species are observed to be monomeric HBD3 and interestingly a shift in the charge state distribution of the monomer to lower charges is also observed, suggesting Fx is promoting the formation of compact monomeric species or that Fx preferentially binds the more extended higher charged species in solution resulting in a subsequent reduction in their relative intensity.

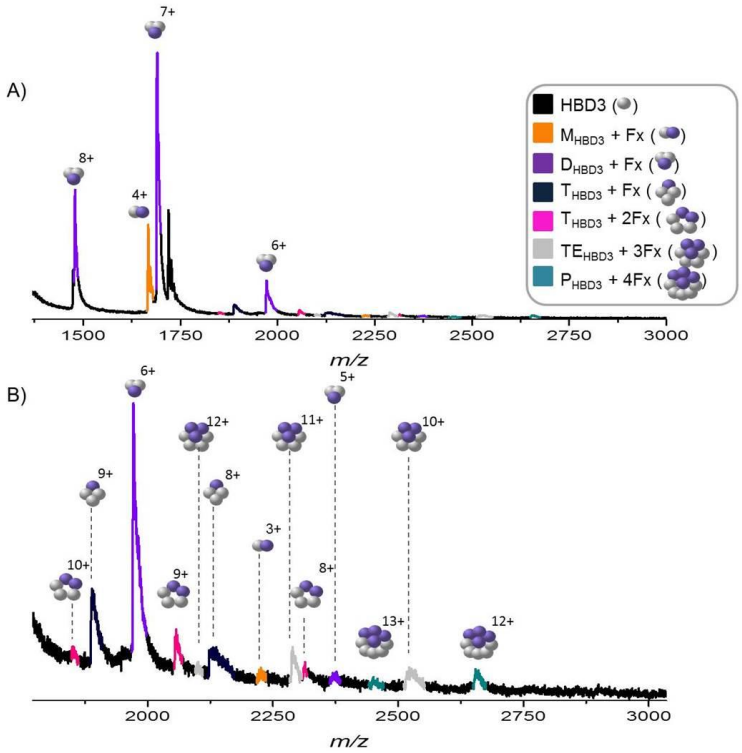

Figure 6.8: Enlarged regions of the mass spectrum (shown in Figure 6.7C) obtained for the HBD3 plus Fx at a 2:1 ratio (50:25 µM) A) *m/z* 1300-3000 B) *m/z* 1800-3000.

A number of additional low intensity peaks are present in the spectrum, which are assigned as HBD3 plus Fx complexes (Figure 6.7C and Figure 6.8). The most intense complex species here is dimeric HBD3 (D_{HBD3}) binding a single Fx chain (D_{HBD3}+Fx), similar to the observations made for HBD2 (Figure 6.2B). The 6+ and 7+ species were isolated and subjected to CID. The first fragmentation pathway was again found to be loss of SO_3 from Fx as opposed to complex dissociation (Appendix 5 Figure A5.6), this highlights the stability of the complexes. For the 7+ species more extensive complex dissociation was also observed but at lower intensity than Fx fragmentation, consistent with the findings for D_{HBD2}+Fx and supports specific, stable complex formation.

Trimeric (T_{HBD3}), tetrameric (TE_{HBD3}) and pentameric (P_{HBD3}) HBD3 are also observed in complex with Fx, where the trimer binds one (T_{HBD3}+Fx) or two (T_{HBD3}+2Fx) Fx chains, tetrameric HBD3 binds three Fx chains (TE_{HBD3}+3Fx) and pentameric HBD3 binds four Fx chains (P_{HBD3}+4Fx). The expected and observed masses of all Fx bound species are in agreement and can be found in Appendix 5 Table A5.5. Fx is clearly promoting aggregation of HBD3 leading to a range of oligomeric species; however, interestingly no species of unbound trimeric or tetrameric HBD3 are observed suggesting under these conditions higher order oligomers are only stabilised by binding to Fx. The sequence of HBD3 does not contain the 'BBXB' amino acid GAG binding motif, suggesting that HBD3 oligomerisation may promote GAG binding through the creation of basic patches to which Fx can bind. The presence of M_{HBD3}+Fx and high GAG content of the oligomers, however, suggests that this motif is not essential for defensin GAG binding and the higher order oligomers are a product of extensive, favourable aggregation which allows more species, potentially on pathway to, or fragments of, higher order aggregates- to be observed.

6.3.2.2 Conformations of HBD3:Fx complexes: insights from DT IM-MS

In samples prepared at a 2:1 HBD3:Fx ratio a wide range of oligomers is detected and hence DT IM-MS studies were performed on these samples allowing CCS to be obtained for a range of species. For HBD3 only unbound monomeric species were observed both in the presence and the absence of Fx. The CCS for monomeric HBD3 were determined in the presence of Fx over the charge state range 4+ to 6+ and found to be ~746-796 $Å^2$ (Appendix 5 Table A5.6). The small CCS range is attributed to the structural rigidity imposed by the three disulfide bonds. The slightly larger size in comparison to HBD2 is due to the longer sequence of HBD3, four amino acids longer than HBD2. CCS were determined for a single repeat of HBD3 in 20 mM AmAc and monomeric species lie within experimental error of those determined in the presence of Fx demonstrating that Fx does not alter the conformation

of unbound HBD3 (Appendix 5 Table A5.6) and in excellent agreement with previously reported values[28].

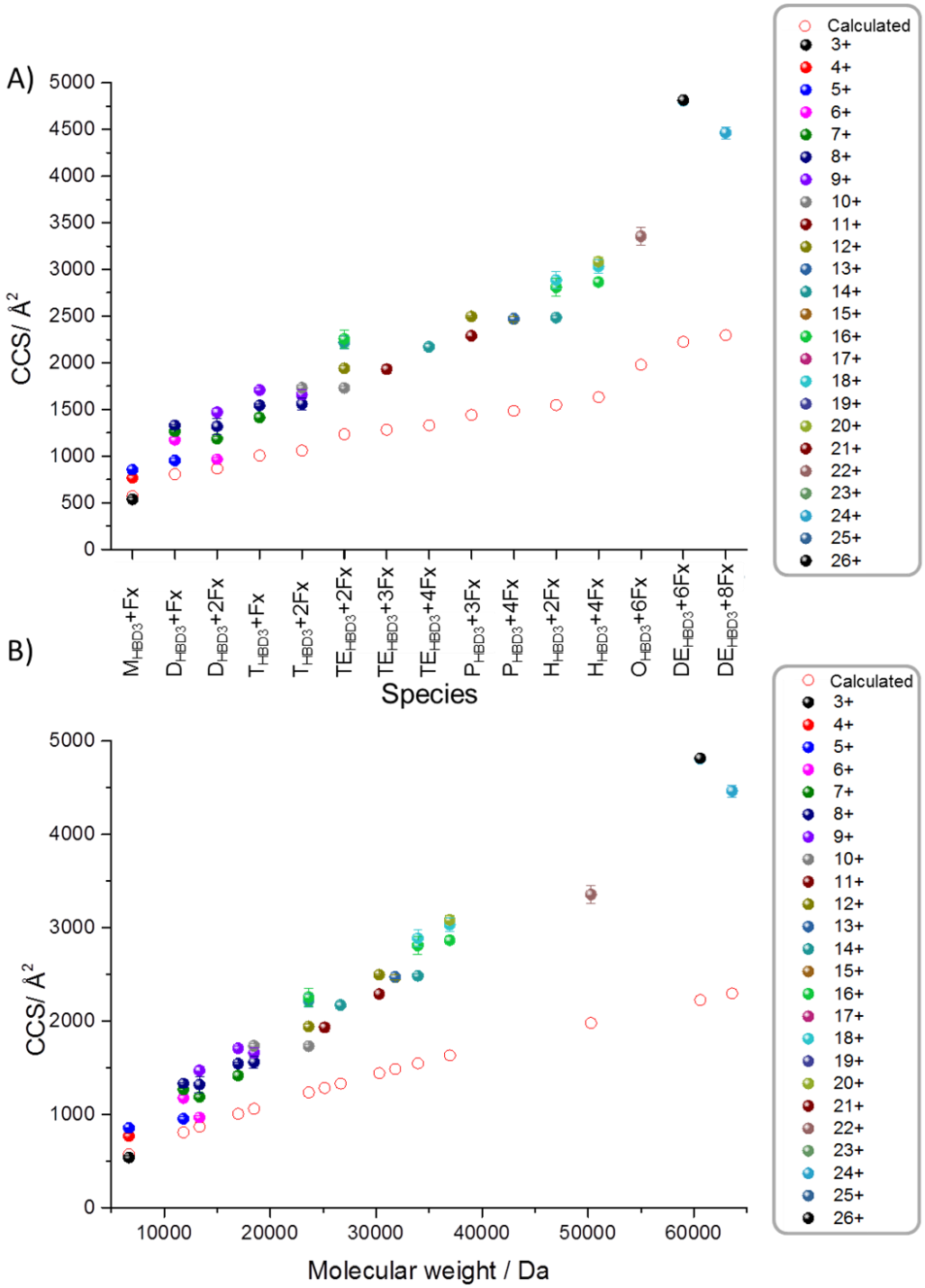

Figure 6.9: Average CCS of Fx bound HBD3 species, from three repeats, error bars represent the standard deviation between these values. Where, monomeric (M), dimeric (D), trimeric (T), tetrameric (TE) pentameric (P), hexameric (H), octameric (O) and decameric (DE) HBD3 species are identified by letter codes. A) species versus CCS B) molecular weight versus CCS. Hollow red symbols represent the CCS predicted for a spherical complex of constant protein density, according to the relationship shown in Equations 6.1.

Fx-bound species were observed in a range of oligomeric orders composed of up to ten HBD3 units and up to six chains of Fx, a number of which were observed as *m/z* coincident species and could only be distinguished using DT IM-MS. The presence and conformations of such a range of GAG-bound peptide oligomers have not to-date been reported, possibly due in part to the problems associated with the extensive and rapid aggregation. Determination of the conformations of these species and their implication in growth mechanisms could provide insight into the nature of their formation. DT IM-MS allows these species to be separated and also their CCS to be determined (Figure 6.9 and Appendix 5 Table A5.7). Due to the range of species observed, viewing the data in a plot of charge state versus CCS, is not optimum for observing trends (Appendix 5 Figure A5.7). Therefore, the data is shown as CCS versus species (Figure 6.9A) and a similar plot of CCS versus molecular weight (Figure 6.9B). Due to the range of oligomeric species observed and the computationally expensive nature of the MD simulations, no theoretical comparisons are made here.

Considering D_{HBD2}+Fx and D_{HBD2}+2Fx, a compaction in CCS is observed moving from the species binding one Fx chain to that binding two at the same charge state (Figure 6.9A; pink, green and navy spheres). This compaction upon binding is interesting as the three disulfide bonds in HBD3 would limit non-specific collapse, which suggests these species are undergoing a conformational constriction or compaction upon binding of additional Fx chains and suggesting there is specificity in their formation. A similar observation is made for the 9+ trimeric species; however, at all other oligomeric orders the same charge state is not observed upon increasing the number of chains of Fx bound so similar comparisons cannot be made here.

The range of CCS covered is expansive, with CCS increasing by over eight times from the M_{HBD3}+Fx species to decameric HBD3 binding six Fx chains (DE_{HBD3}+6Fx). Interestingly, growth appears somewhat linear over the range of species observed here, with no major plateau region, significant jump or compression in CCS over the full range of species. There are a number of growth models which have previously reported to fit to CCS data including: the isotropic growth model, single strand fibril growth and steric zipper fibril growth[55]. Fibril growth mechanisms are not appropriate here, as TEM studies show the resulting species are not fibrillar in nature (see section 6.3.2.3). The isotropic growth model states that the collision cross section (Ω) is related to the oligomeric state (n) by $\Omega = \Omega_m n^{2/3}$, where Ω_m is the monomer cross section. The difficulty with fitting this model to the data presented here is that it does not account for hetero-oligomers and hence would be inappropriate for this data

where growth occurs through binding of two different species. It is possible, instead of considering the monomer as the initial growth state, to consider M_{HBD3}+Fx, however, this growth model would then only be appropriate for species in which even numbers of HBD3 and Fx are binding, which is only the case for three of the observed species namely; monomer (M_{HBD3}+Fx), dimer (D_{HBD3}+2Fx) and tetramer (TE_{HBD3}+4Fx). Hence, the isotropic growth model is not considered here.

Benesch *et al*[56] have provided a relationship through which the CCS can be estimated, assuming species are adopting a perfect spherical conformation as shown in Equation 6.1, where Ω is the CCS, *m* is the mass, ρ is the density and r_g is the radius of the buffer gas (which for helium is 1.4 Å). Numerous estimates of protein density have been reported in the literature[57,58] and here a value of 0.90 Da $Å^{-3}$ is used, which is based on small globular proteins[58]. CCS of all Fx bound species were estimated in this way (Appendix 5 Table A5.7) and are shown plotted as hollow red circles in Figure 6.9. These spherical CCS are much lower than the experimentally observed species as expected, consistent with the observed oligomers existing as specific complexes undergoing a form of linear growth and not existing as purely spherical species of constant density, species which would be unlikely to form. Oligomers formed, therefore, are non-spherical and are, with respect to size, between globular and linear species.

$$\Omega = \pi\left(\sqrt[3]{\frac{3m}{4\pi p}} + r_g\right)^2 \qquad [6.1]$$

6.3.2.3 Assessing aggregation: TEM studies

From MS it is clear that extensive aggregation occurs when HBD3 and Fx are mixed at a 1:1 ratio, a finding which is further accentuated through TEM studies (Figure 6.6A). Images obtained from TEM show a number of large globular aggregates, similar in appearance to those observed for HBD2 in the presence of Fx. HBD3 aggregate species, however, appear larger and aggregation is much more extensive, covering a larger proportion of the TEM grid. The extent of aggregation occurring is noteworthy considering samples were incubated for only 30 minutes prior to grid preparation. The aggregates form large clusters with many species overlapping causing the morphology of the species to be difficult to discern.

In order to further probe the aggregation occurring for HBD3 upon addition of Fx a further sample was studied at a 2:1 HBD3:Fx ratio (Figure 6.10B), as MS studies highlighted that at this ratio aggregation can be slowed enabling a greater range of oligomers to be observed. TEM images of the 2:1 mixture of HBD3 and Fx show similar species to those formed from

the 1:1 mixture of HBD3 plus Fx, large globular aggregate clusters of typically of μm length. At this 2:1 ratio, as aggregation is slowed, smaller clusters of species are also observable in the background. These background species appear of a similar globular nature, however, 'networks' of interconnecting species can be distinguished which presumably go on to form the large globular 'networks' present in the more aggregated sample. As a control HBD3 in the absence of Fx was studied and no evidence for higher order aggregation was observed, indicating this is an Fx induced effect (Appendix 5 Figure A5.8). Clearly Fx induces aggregation in both HBD2 and HBD3, with both defensins forming aggregates of a globular morphology.

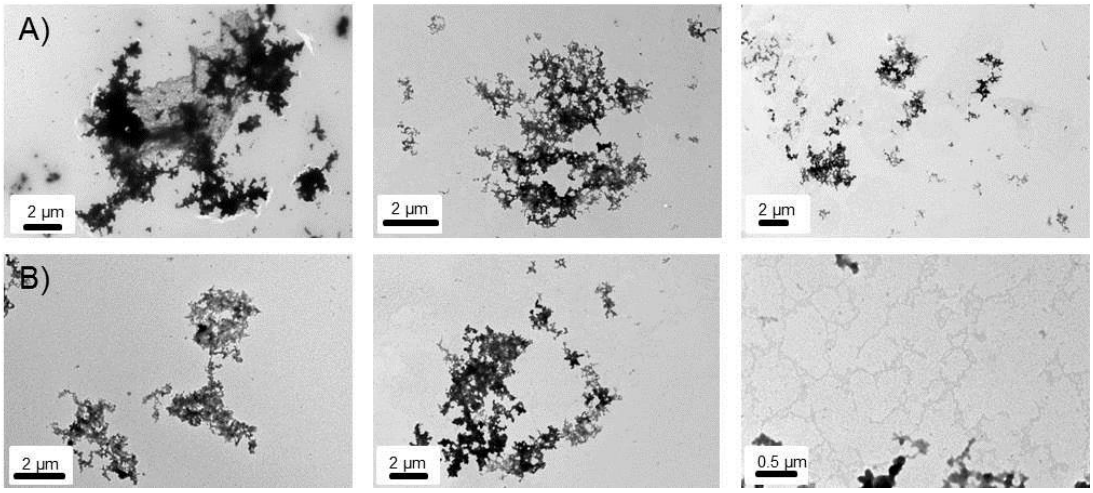

Figure 6.10: Representative TEM images obtained for A) 1:1 HBD3 plus Fx (50 μM) and B) 2:1 HBD3 plus Fx (50 μM HBD3 concentration).

6.3.3 Investigating HBD2:HBD3 interactions

6.3.3.1 Stoichiometry of binding: MS studies

The results in the previous sections highlight that both HBD2 and HBD3 can bind Fx, here the ability of these two β-defensins to interact with each other is investigated. The need for such a study is based on the apparent redundancy of the immune system, with many proteins interacting with the same receptor; for example HBDs can activate the G protein coupled receptor CCR6[19,59], which is also activated by the chemokine CCL20[60]. Recent studies have shown that chemokines can form heterodimers which can also bind GAGs, challenging the previously considered redundancy of these systems[37,38,44] and hence equivalent studies are required for β-defensins.

Initial studies were performed on a 1:1 solution of HBD2 and HBD3 and the resulting spectrum is dominated by the unbound peptides, mostly in a monomeric form (Figure 6.11).

Additionally, HBD2 forms dimers and trimers and HBD3 is also observed in a dimeric form, both of which are consistent with findings for the individual peptides (Figure 6.2 and Figure 6.7). However, in the case of HBD3 the dimeric species are found to be more intense in this mixed sample.

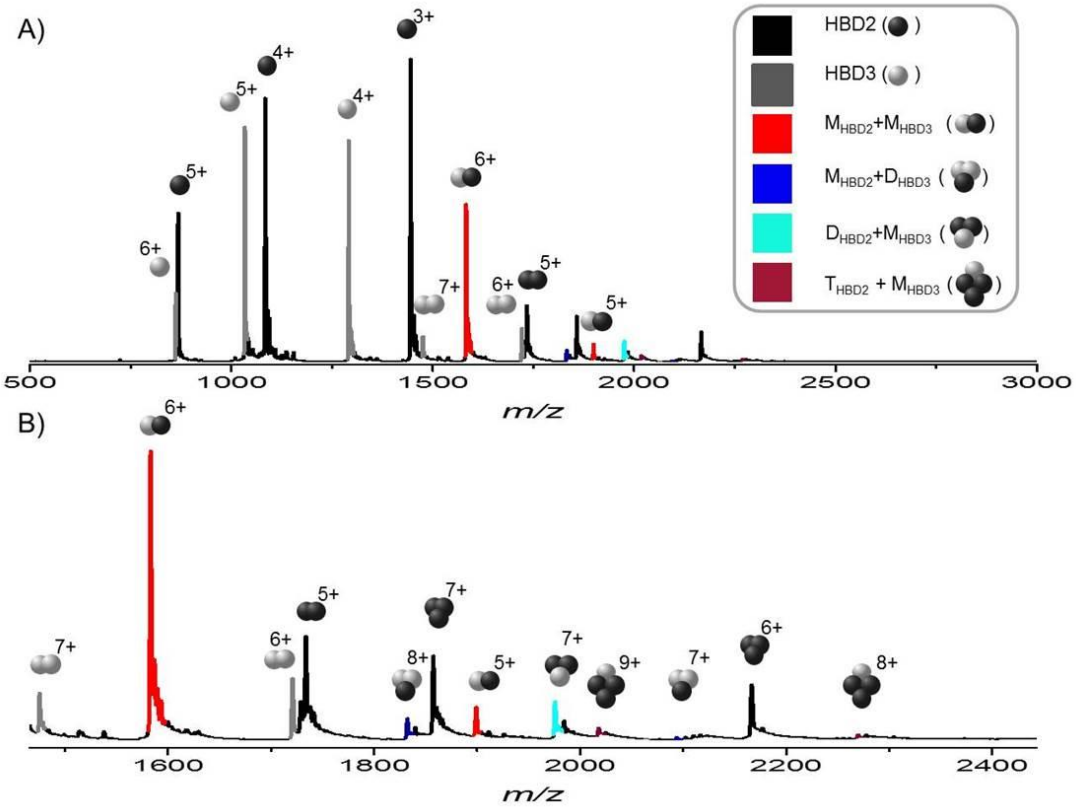

Figure 6.11: Mass spectrum of HBD3 plus HBD3 prepared at a 1:1 ratio at 50 µM concentration A) *m/z* 500-3000 B) Enlargement of *m/z* 1465-2440. Peptide monomer (M), dimer (D) and trimer are described by letter codes.

A number of additional peaks due to HBD2+HBD3 complexes are also observed (Figure 6.11). The heterodimeric complex ($M_{HBD2}+M_{HBD3}$) is present at high intensity demonstrating that the formation of this complex is favourable. Furthermore, a heterotrimer composed of two HBD2 plus one HBD3 ($D_{HBD2}+M_{HBD3}$) and heterotetramers composed of three HBD2 plus one HBD3 ($T_{HBD2}+M_{HBD3}$) are formed. It is proposed that for both these cases HBD3 is binding to the HBD2 dimers and trimers already present in solution. A corresponding substantial decrease in relative intensity of the unbound HBD2 dimers and trimers is not observed, suggesting that the original amount of unbound dimer and trimer are restored by a shift in equilibrium, between the monomeric species and these higher order oligomers, according to Le Chatelier's principle. Additionally, heterotrimers are detected at two charge states which comprise one HBD2 plus two HBD3 ($M_{HBD2}+D_{HBD3}$). These species could be

formed *via* the HBD3 dimers binding monomeric HBD2 and hence could explain the promoted formation and higher intensity of HBD3 dimers here. The expected and observed masses of all heterodefensin complexes are in good agreement and can be found in Appendix 5 Table A5.8.

Next the ability of these heterodefensin complexes to bind Fx was studied. Here a 1:1:1 sample of HBD2 and HBD3 plus Fx was prepared with each component present at 25 µM. The most intense peaks in this spectrum (Figure 6.12) are again attributed to the individual peptides, however, in this case HBD2 is significantly more intense due to its apparent lower binding affinity for Fx, forming fewer observable species in comparison to HBD3. HBD3 binds Fx in a range of oligomeric orders from monomer to pentamer, consistent with the findings for the 2:1 HBD3 plus Fx sample studied earlier (Figure 6.7). Significantly, only one low intensity HBD2+Fx complex is observed, trimeric HBD2 binding a single Fx chain (T_{HBD2}+Fx), in contrast to the high intensity dimeric Fx-bound species observed in the absence of HBD3 (Figure 6.2), suggesting HBD3 out-competes formation of this species. The intensity of the M_{HBD2}+M_{HBD3} heterodimer is also significantly reduced, attributed again to the rapid and favourable binding of Fx to HBD3.

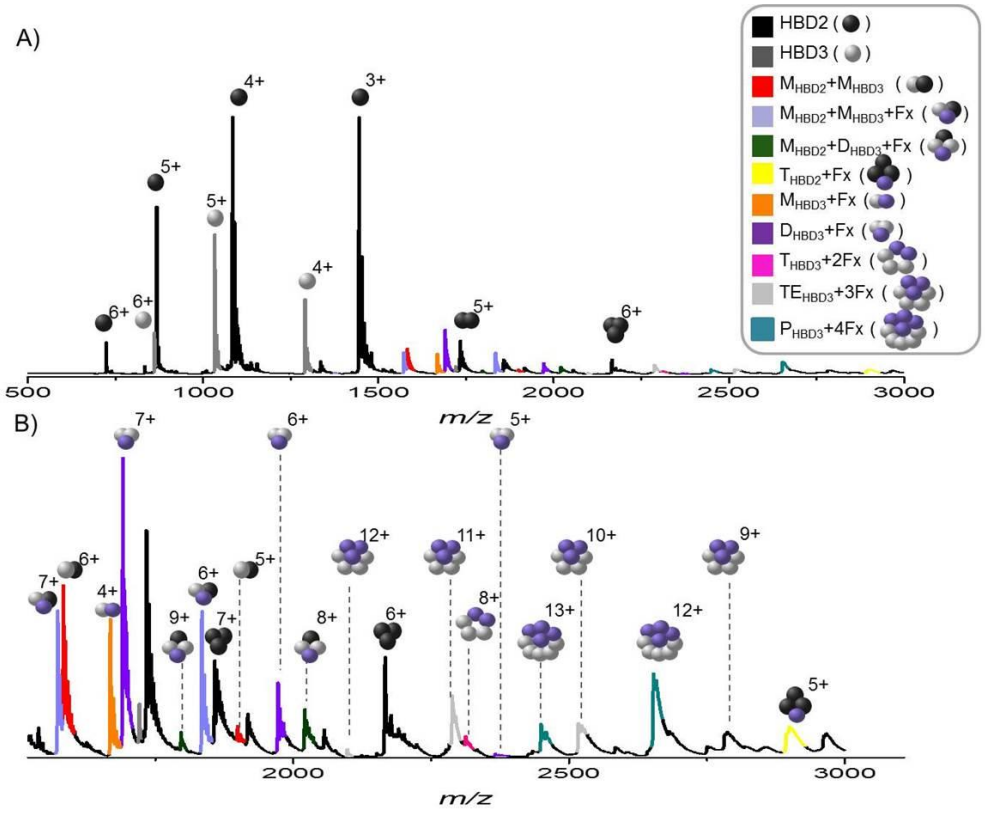

Figure 6.12: Mass spectrum of HBD3 plus HBD3 plus Fx prepared at a 2:2:1 ratio at 25 µM peptide concentration A) *m/z* 500-3000 B) Enlargement of *m/z* 1500-3000. Peptide monomeric (M), dimeric (D), trimeric (T), tetrameric (TE) and pentameric (P) species are identified by letter codes.

Interestingly, a number of HBD2+HBD3+Fx species are present, namely a M_{HBD2}+M_{HBD3}+Fx complex observed in two charge states (6+ and 7+) and a M_{HBD2}+D_{HBD3}+Fx complex detected in the 8+ and 9+ charge states. Both the M_{HBD2}+M_{HBD3} and M_{HBD2}+D_{HBD3} complexes were observed in absence of Fx; however, the charge state distribution is lower by a single charge in the unbound form. The higher charge states of the Fx-bound species cannot be attributed to the Fx, and instead suggests that these complexes are formed from a higher charged monomer of either HBD2 or HBD3 in the presence of Fx. An alternative explanation is that the Fx-bound heterodefensin complexes have a different binding interface than the heterodefensin complexes themselves and therefore the surface, solvent exposed residues are different here leading to more solvent accessible chargeable residues, a theory which can be probed further using DT IM-MS and is discussed in detail below.

Control MS experiments were performed with HBD2 and bovine pancreatic trypsin inhibitor (BPTI), to discern the specificity of binding of HBD2 and HBD3. BPTI was chosen due to its similar tertiary fold to the β-defensins, comprising two β-strands and two helices constrained by three disulfide bonds. In this case in a 1:1 mixture no complexes between HBD2 and BPTI form, suggesting the binding between HBD2 and HBD3 is specific (Appendix 5 Figure A5.9A). Low intensity 1:1:1 (HBD2:BPTI:Fx) complexes were observed in the presence of Fx, however, highlighting Fx can promote complex formation (Appendix 5 Figure A5.9B).

The presence of high intensity heterodefensin complexes is, therefore, significant. The ability of β-defensins to form complexes between different members of this subclass has not previously been investigated and the determination that these complexes can go on to bind Fx proposes they could have a functional role within GAG binding and chemotaxis.

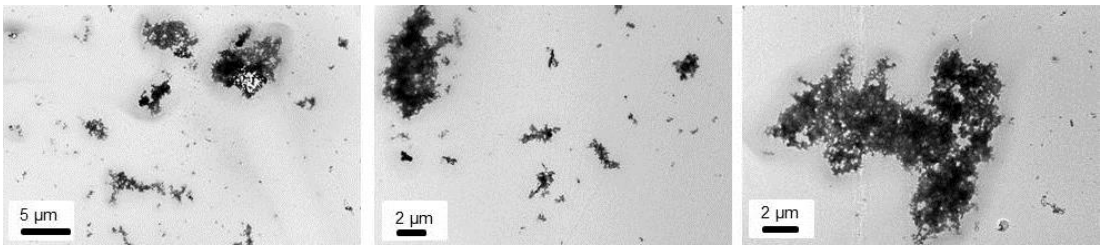

Figure 6.13: Representative TEM images obtained for a 1:1:1 mixture of HBD2 plus HBD3 plus Fx.

It is important to note that upon addition of Fx, the mixed defensin sample was seen to aggregate within the n-ESI glass capillary, an observation not made with the HBD2 plus HBD3 in the absence of Fx. Therefore, as with the individual peptides plus Fx, TEM studies

were performed (Figure 6.13). From the TEM images it is clear the aggregation occurring in the mixed peptide plus Fx sample is similar to that in the individual peptide samples in the presence of Fx with networks of large globular aggregates forming.

6.3.3.2 Conformations of heterodefensin complexes: insights from DT IM-MS

In order to further probe the interactions of HBD2 and HBD3 DT IM-MS experiments were performed. CCS were first determined for the heterodefensin complexes and compared to the individual monomers and dimers (Figure 6.14 and Appendix 5 Table A5.9). CCS of the individual monomers and dimers plotted in Figure 6.14 are those determined in the individual peptide plus Fx sample (reported in sections 6.3.1.2 and 6.3.2.2). CCS for the monomeric species of HBD2 and HBD3 were also determined from the HBD2 plus HBD3 mixed sample and remain the same within experimental error (Appendix 5 Table A5.10), signifying the presence of the other peptide does not influence the monomeric conformations.

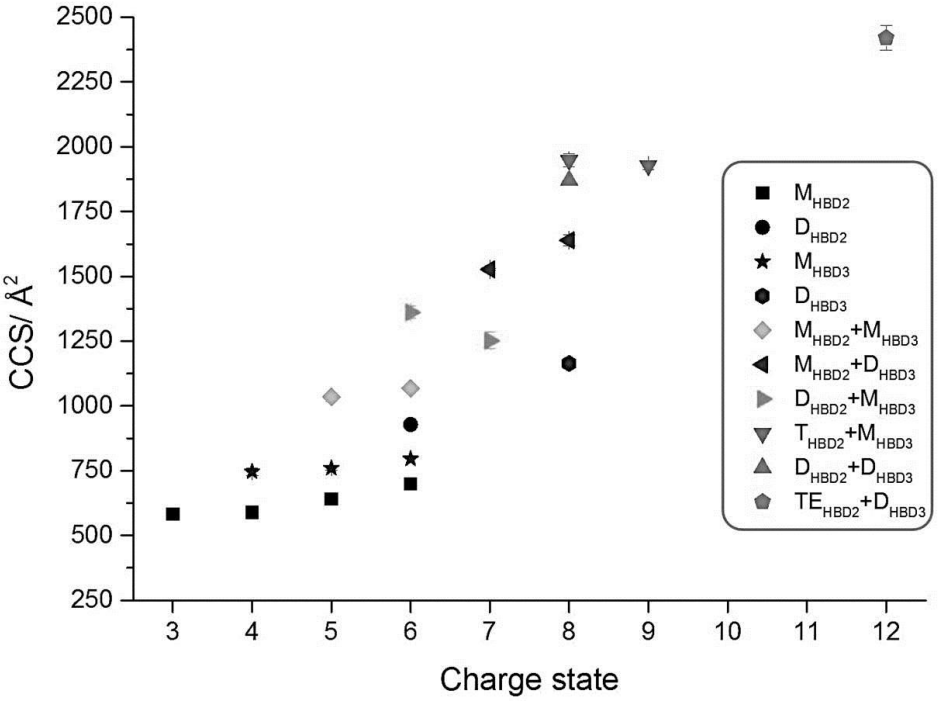

Figure 6.14: Average CCS of HBD2, HBD3 and the heterodefensin complexes, calculated from three repeats, error bars represent the standard deviation between these values. HBD2 and HBD3 CCS were as determined from individual peptide solutions. Where peptide monomeric (M), dimeric (D), trimeric (T) and tetrameric (TE) species are identified by letter codes.

Considering next the heterodimers (M_{HBD2}+ M_{HBD3}), CCS were determined for the 5+ and 6+ complexes and were found to increase marginally with charge, lying between those of the homodimers, consistent with stable, specific complexes. CCS were determined for the two

different heterotrimeric complexes, namely $D_{HBD2}+M_{HBD3}$ and $M_{HBD2}+D_{HBD3}$. Interestingly, the trimers $D_{HBD2}+M_{HBD3}$ are significantly smaller than the $M_{HBD2}+D_{HBD3}$ trimers, furthermore these species are seen to become more compact with increasing charge in contrast to $M_{HBD2}+D_{HBD3}$ which becomes more extended with increasing charge. The substantial differences in the CCS for these two families of trimeric species suggests that they are different in nature, with an excess of HBD2 promoting tighter packing and more compact CCS. Interestingly, the $T_{HBD2}+M_{HBD3}$ species also become more compact with increasing charge state suggesting this is characteristic of complexes containing more HBD2 than HBD3; however, in this case the compaction is much lower, lying within experimental error. In addition the $D_{HBD2}+D_{HBD3}$ complex is very similar in CCS to that of the $M_{HBD2}+D_{HBD3}$ complex at the same charge state, proposing similar complexes are formed here.

Clearly, the conformations of the complexes are diverse; however, a number of trends can be recognised here. When HBD2 is in excess, in the complex the species presents with small CCS and undergoes compaction with increasing charge. In contrast, when HBD3 is in excess species are generally larger. This difference in CCS is greater than that found for the individual peptides and is therefore not attributed to the longer chain length of HBD3 and instead inferred to be due to differences in the complex conformations and resulting packing of subunits.

Upon addition of Fx, the conformations of the heterodefensin complexes observed in the absence and presence of Fx were considered, namely the heterodimers at charge states 5+ and 6+, and found to lie within experimental error (Appendix 5 Table A5.9). The conformations adopted by the GAG-bound heterodefensin complexes were then determined (Figure 6.15 and Appendix 5 Table A5.11). Considering the $M_{HBD2}+M_{HBD3}+Fx$ complex first, it is clear two conformational families are present. The more compact conformational family presents over a narrow CCS range (~1134-1301 $Å^2$) over the charge state range 5+ to 8+, whilst the more extended conformational family is observed only at charge states 7+ and 8+ over a CCS range of ~1429-1542 $Å^2$. CCS of the compact family are marginally larger than the $M_{HBD2}+M_{HBD3}$ complex, suggesting the conformation is not significantly altered upon binding, whilst the more extended family is considerably larger than the unbound species suggesting the conformations these species are adopting are significantly different than those of the unbound heterodimers, and suggests these species may have a different interaction interface. The presence of two distinct conformational families suggests there may also be two discrete binding sites for this heterodimeric complex, which, considering the nature of the subunits, both of which can bind Fx, is probable.

CCS for M_{HBD2}+D_{HBD3}+Fx were determined for the 9+ charge state and this species was found to exist in two conformational families, the more compact of which is similar to that of the more extended family of M_{HBD2}+M_{HBD3}+Fx. Significantly, the CCS of this Fx-bound trimer was observed to be smaller than that of its unbound counterpart at the 8+ charge state but similar in CCS to that of the $[M_{HBD2}+D_{HBD3}+7H]^{7+}$ complex. This lower CCS suggests that Fx confers a conformational tightening, consistent with stable specific binding. The second conformation is much more extended and illustrates a significantly different complex is formed here, perhaps due to Fx binding to a different, conceivably a less specific, binding site. Additionally, the trimerisation interface may be different for this extended complex.

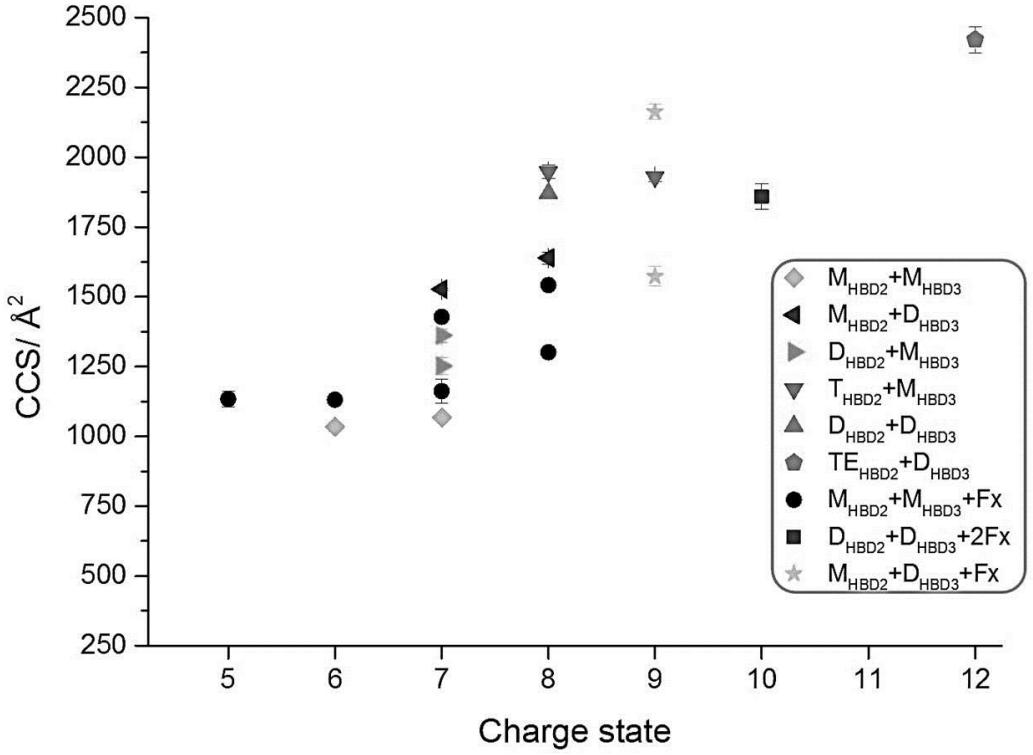

Figure 6.15: Average CCS of heterodefensin complexes unbound and bound to Fx, calculated from three repeats, error bars represent the standard deviation between these values. Fx unbound heterodefensin complex CCS were determined in the absence of Fx. Where peptide monomeric (M), dimeric (D), trimeric (T) and tetrameric (TE) species are identified by letter codes.

The final Fx-bound heterodefensin complex detected was $[D_{HBD2}+D_{HBD3}+2Fx+10H]^{10+}$, which presents at the same *m/z* as $[M_{HBD2}+M_{HBD3}+Fx+5H]^{5+}$. The CCS of this hexameric species is within error of the D_{HBD2}+D_{HBD3} complex observed at its single charge state (8+). As only one charge state is observed for this species in both its Fx-bound and unbound form it is difficult to discern any trends; however, the similar CCS suggest that binding of Fx may be specific. These results demonstrate that Fx can specifically bind a number of heterodefensin complexes, reinforcing the possibility of a functional role. Further *in vitro* and *in vivo* studies would be required to elucidate the functional relevance of these species.

6.4 Conclusions

β-defensins share a number of structural and functional features with the chemokine superfamily of proteins. The studies presented here were designed to further investigate these similarities through consideration of GAG binding. Both HBD2 and HBD3 bind the model GAG fondparinux forming complexes of a similar nature to those of chemokines, particularly in the case of HBD2.

HBD2 binds in a controlled, specific manner, exclusively binding Fx in a dimeric form, consistent with previous studies which have shown a number of chemokines bind GAGs in dimeric forms. However, preferential formation of GAG-bound monomers or dimers is thought to depend upon the chain length of the GAG[31,33,61]. Hence it would be of interest to study HBD2 interactions with GAGs of varying chain length by MS and other biophysical techniques. The dimeric HBD2 plus Fx complexes were studied using a variety of biophysical techniques. DT IM-MS, in combination with detailed MD simulations, provided a likely complex candidate structure. The stability of this complex is illustrated through CID and ECD studies in which limited fragmentation and dissociation occurs. Finally, ITC shows this binding process is complex, dominated by endothermic processes at low molar ratio and exothermic processes at high molar ratios. There are several possible processes which could be occurring here and therefore many conceivable explanations for this behaviour; however, to better understand these processes more robust data, including numerous repeats, would be required which due to limited sample were not possible here.

HBD3 aggregates extensively upon addition of Fx, resulting in a range of oligomeric species (monomer to decamer) binding up to eight chains of Fx. Such a wide range of GAG-bound oligomers has not to-date been reported, which could in part be attributed to the difficulty in studying such systems due to extensive aggregation. This is exemplified for HBD3 through the TEM studies, in which large ‘networks’ of globular aggregates are observed, similar in appearance to those observed for WT 1-72 Ltn plus Fx (Chapter 5 section 5.3.2), the Ltn construct with the most similar structure to the HBDs. The possible functional relevance of aggregating species is of interest; formation could overcome the effects of vascular flow, enabling a concentration gradient to be established and it is conceivable that these large aggregate species may be more resistant to digestion. The binding of β-defensins with GAGs could have implications for the chemotactic activity of these peptides. The results suggest that, as with chemokines, interactions with cell surface GAGs may enable a concentration gradient along which cells can migrate *in vivo*. Furthermore, previous studies have demonstrated that GAG binding can inhibit the antimicrobial activity of defensins[25,26], which

bacteria could exploit through the induction of GAGs[62], decreasing the efficacy of host defence. Clearly defensin GAG binding is a complex, biologically important process and better understanding these interactions at both fundamental *in vitro* and *in vivo* levels with respect to both antimicrobial and chemotactic activities is essential.

Furthermore, the interactions between HBD2 and HBD3 were studied. HBD2 and HBD3 are shown here to interact forming a range of heterodefensin complexes from dimers to hexamers. This finding was not replicated in experiments considering HBD2 and BPTI, suggesting heterodefensin complexes are specific. The conformations adopted by all heterodefensin complexes were studied using DT IM-MS and a remarkable trend was observed whereby the species in which HBD2 is in excess are more compact than HBD3 counterparts, and become more compact with increasing charge, indicating that HBD2 promotes the formation of tighter complexes. The ability of heterodefensin complexes to bind Fx was then considered, the resulting data show a clear preference for HBD3:Fx binding, in a range of oligomeric orders, as opposed to HBD2 or heterodefensin complexes binding Fx, further illustrating the high propensity of HBD3 for GAG binding. Despite this, dimeric, trimeric and tetrameric heterodefensin complexes were observed to bind GAGs with CCS of all species suggesting that stable, specific complexes are forming.

The ability of different members of the β-defensin subclass to interact with each other, and to go on to interact with other biomolecules, has not to-date been reported. The observation of such species at high intensity, forming specific complexes is significant and suggests there could be a functional role for such complexes. It is conceivable that hetero-oligomerisation could play a role in differentiating between and modulating the activity of specific defensins. In-depth studies would be required, however, to determine any effects, to conclude if these species can form *in vivo* and if the functional response is different for the heterodefensin mixtures as opposed to the individual samples, an effect which has been previously observed for heterochemokines[37,42]. Although it is thought that monomers bind the GPCR, dimerisation[23], and possibly heterodimerisation, is thought to be important for *in vivo* function, possibly to overcome effects of vascular flow or to increase concentration at the receptor.

The processes of GAG binding and the interactions between different members of the β-defensin subfamily are intricate and complex. As cells exist in and migrate through complex environments the consideration and understanding of possible interactions is vital and may contribute to the better understanding of the sophisticated immunological system. Through the combination of techniques presented here insight can be gained, enabling the

species to be identified and the conformations of both the initial complexes and the highly aggregated samples to be studied.

6.5 References

(1) Hoover, D. M.; Rajashankar, K. R.; Blumenthal, R.; Puri, A.; Oppenheim, J. J.; Chertov, O.; Lubkowski, J., *J. Biol. Chem.*, **2000**, *275*, 32911-32918.
(2) Sawai, M. V.; Jia, H. P.; Liu, L.; Aseyev, V.; Wiencek, J. M.; McCray, P. B.; Ganz, T.; Kearney, W. R.; Tack, B. F., *Biochemistry*, **2001**, *40*, 3810-3816.
(3) Schibli, D. J.; Hunter, H. N.; Aseyev, V.; Starner, T. D.; Wiencek, J. M.; McCray, P. B., Jr.; Tack, B. F.; Vogel, H. J., *J. Biol. Chem.*, **2002**, *277*, 8279-8289.
(4) Pazgier, M.; Hoover, D.; Yang, D.; Lu, W.; Lubkowski, J., *Cell. Mol. Life Sci.*, **2006**, *63*, 1294-1313.
(5) Fernandez, E. J.; Lolis, E., *Annu. Rev. Pharmacol. Toxicol.*, **2002**, *42*, 469-499.
(6) Hoover, D. M.; Boulègue, C.; Yang, D.; Oppenheim, J. J.; Tucker, K.; Lu, W.; Lubkowski, J., *J. Biol. Chem.*, **2002**, *277*, 37647-37654.
(7) Lehrer, R. I., *Nat. Rev. Microbiol.*, **2004**, *2*, 727-738.
(8) Selsted, M. E.; Ouellette, A. J., *Nat. Immunol.*, **2005**, *6*, 551-557.
(9) Hamel, D. J.; Sielaff, I.; Proudfoot, A. E.; Handel, T. M., *Methods Enzymol.*, **2009**, *461*, 71-102.
(10) Krijgsveld, J.; Zaat, S. A.; Meeldijk, J.; van Veelen, P. A.; Fang, G.; Poolman, B.; Brandt, E.; Ehlert, J. E.; Kuijpers, A. J.; Engbers, G. H., *J. Biol. Chem.*, **2000**, *275*, 20374-20381.
(11) Yang, D.; Chen, Q.; Hoover, D. M.; Staley, P.; Tucker, K. D.; Lubkowski, J.; Oppenheim, J. J., *J. Leukoc. Biol.*, **2003**, *74*, 448-455.
(12) Yung, S. C.; Parenti, D.; Murphy, P. M., *J. Biol. Chem.*, **2011**, *286*, 5069-5077.
(13) Burkhardt, A. M.; Tai, K. P.; Flores-Guiterrez, J. P.; Vilches-Cisneros, N.; Kamdar, K.; Barbosa-Quintana, O.; Valle-Rios, R.; Hevezi, P. A.; Zuñiga, J.; Selman, M., *J. Immunol.*, **2012**, *188*, 6399-6406.
(14) Arias, M.; Zaat, S. A.; Vogel, H. J. In *Antimicrobial Peptides and Innate Immunity*; Springer: 2013, p 183-218.
(15) Esche, C.; Stellato, C.; Beck, L. A., *J. Investig. Dermatol.*, **2005**, *125*, 615-628.
(16) Allen, S. J.; Crown, S. E.; Handel, T. M., *Annu. Rev. Immunol.*, **2007**, *25*, 787-820.
(17) Baggiolini, M.; Dewald, B.; Moser, B., *Annu. Rev. Immunol.*, **1997**, *15*, 675-705.
(18) Röhrl, J.; Yang, D.; Oppenheim, J. J.; Hehlgans, T., *J. Immunol.*, **2010**, *184*, 6688-6694.
(19) Yang, D.; Chertov, O.; Bykovskaia, S.; Chen, Q.; Buffo, M.; Shogan, J.; Anderson, M.; Schröder, J.; Wang, J.; Howard, O., *Science*, **1999**, *286*, 525-528.
(20) Soruri, A.; Grigat, J.; Forssmann, U.; Riggert, J.; Zwirner, J., *Eur. J. Immunol.*, **2007**, *37*, 2474-2486.
(21) Oppenheim, J.; Biragyn, A.; Kwak, L.; Yang, D., *Ann. Rheum. Dis.*, **2003**, *62*, ii17-ii21.
(22) Handel, T. M.; Johnson, Z.; Crown, S. E.; Lau, E. K.; Proudfoot, A. E., *Annu. Rev. Biochem.*, **2005**, *74*, 385-410.
(23) Proudfoot, A. E. I.; Handel, T. M.; Johnson, Z.; Lau, E. K.; LiWang, P.; Clark-Lewis, I.; Borlat, F.; Wells, T. N. C.; Kosco-Vilbois, M. H., *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, *100*, 1885-1890.
(24) Andersson, E.; Rydengård, V.; Sonesson, A.; Mörgelin, M.; Björck, L.; Schmidtchen, A., *Eur. J. Biochem.*, **2004**, *271*, 1219-1226.
(25) Schmidtchen, A.; Frick, I. M.; Björck, L., *Mol. Microbiol.*, **2001**, *39*, 708-713.
(26) McCullough, B.; Kalapothakis, J.; Wutharath, C.; Taylor, K.; Clarke, D. J.; Eastwood, H.; Campopiano, D.; MacMillan, D.; Dorin, J. R.; Barran, P., *Phys. Chem. Chem. Phys.*, **2010**, *12*, 3589-3596.
(27) Rapista, A.; Ding, J.; Benito, B.; Lo, Y.-T.; Neiditch, M. B.; Lu, W.; Chang, T. L., *Retrovirology*, **2011**, *8*, 45.
(28) De Cecco, M., University of Edinburgh, 2011.

(29) Seo, E. S.; Blaum, B. S.; Vargues, T.; De Cecco, M.; Deakin, J. A.; Lyon, M.; Barran, P. E.; Campopiano, D. J.; Uhrin, D., *Biochemistry*, **2010**, *49*, 10486-10495.
(30) Yu, Y.; Sweeney, M. D.; Saad, O. M.; Crown, S. E.; Handel, T. M.; Leary, J. A., *J. Biol. Chem.*, **2005**, *280*, 32200-32208.
(31) Yu, Y.; Sweeney, M. D.; Saad, O. M.; Leary, J. A., *J. Am. Soc. Mass Spectrom.*, **2006**, *17*, 524-535.
(32) Jen, C. H.; Leary, J. A., *Anal. Biochem.*, **2010**, *407*, 134-140.
(33) Seo, Y.; Andaya, A.; Bleiholder, C.; Leary, J. A., *J. Am. Chem. Soc.*, **2013**, *135*, 4325-4332.
(34) Jansma, A.; Handel, T. M.; Hamel, D. J., *Methods Enzymol.*, **2009**, *461*, 31-50.
(35) Tuinstra, R. L.; Peterson, F. C.; Kutlesa, S.; Elgin, E. S.; Kron, M. A.; Volkman, B. F., *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105*, 5057-5062.
(36) Clark-Lewis, I.; Kim, K.; Rajarathnam, K.; Gong, J.-H.; Dewald, B.; Moser, B.; Baggiolini, M.; Sykes, B., *J. Leukoc. Biol.*, **1995**, *57*, 703-711.
(37) Dudek, A. Z.; Nesmelova, I.; Mayo, K.; Verfaillie, C. M.; Pitchford, S.; Slungaard, A., *Blood*, **2003**, *101*, 4687-4694.
(38) Nesmelova, I. V.; Sham, Y.; Dudek, A. Z.; van Eijk, L. I.; Wu, G.; Slungaard, A.; Mortari, F.; Griffioen, A. W.; Mayo, K. H., *J. Biol. Chem.*, **2005**, *280*, 4948-4958.
(39) Weber, C.; Koenen, R. R., *Trends Immunol.*, **2006**, *27*, 268-273.
(40) Tang, Y.-Q.; Yeaman, M. R.; Selsted, M. E., *Infect. Immun.*, **2002**, *70*, 6524-6533.
(41) Nesmelova, I. V.; Sham, Y.; Gao, J.; Mayo, K. H., *J. Biol. Chem.*, **2008**, *283*, 24155-24166.
(42) Paoletti, S.; Petkovic, V.; Sebastiani, S.; Danelon, M. G.; Uguccioni, M.; Gerber, B. O., *Blood*, **2005**, *105*, 3405-3412.
(43) Hoogewerf, A. J.; Kuschert, G. S.; Proudfoot, A. E.; Borlat, F.; Clark-Lewis, I.; Power, C. A.; Wells, T. N., *Biochemistry*, **1997**, *36*, 13570-13578.
(44) Crown, S. E.; Yu, Y.; Sweeney, M. D.; Leary, J. A.; Handel, T. M., *J. Biol. Chem.*, **2006**, *281*, 25438-25446.
(45) Harder, J.; Bartels, J.; Christophers, E.; Schroder, J., *Nature*, **1997**, *387*, 861-861.
(46) Harder, J.; Bartels, J.; Christophers, E.; Schröder, J.-M., *J. Biol. Chem.*, **2001**, *276*, 5707-5713.
(47) Klotman, M. E.; Chang, T. L., *Nat. Rev. Immunol.*, **2006**, *6*, 447-456.
(48) McCullough, B. J.; Kalapothakis, J.; Eastwood, H.; Kemper, P.; MacMillan, D.; Taylor, K.; Dorin, J.; Barran, P. E., *Anal. Chem.*, **2008**, *80*, 6336-6344.
(49) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., *J. Comput. Chem.*, **2009**, *30*, 2785-2791.
(50) Proudfoot, A. E.; Fritchley, S.; Borlat, F.; Shaw, J. P.; Vilbois, F.; Zwahlen, C.; Trkola, A.; Marchant, D.; Clapham, P. R.; Wells, T. N., *J. Biol. Chem.*, **2001**, *276*, 10620-10626.
(51) Iavarone, A. T.; Paech, K.; Williams, E. R., *Anal. Chem.*, **2004**, *76*, 2231-2238.
(52) Iavarone, A. T.; Williams, E. R., *Int. J. Mass. Spectrom.*, **2002**, *219*, 63-72.
(53) Sterling, H.; Kintzer, A.; Feld, G.; Cassou, C.; Krantz, B.; Williams, E., *J. Am. Soc. Mass Spectrom.*, **2012**, *23*, 191-200.
(54) Remko, M.; Van Duijnen, P. T.; Broer, R., *RSC Advances*, **2013**, *3*, 1789-1796.
(55) Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T., *Nat. Chem.*, **2011**, *3*, 172-177.
(56) Benesch, J. L. P.; Ruotolo, B.; Simmons, D. A.; Robinson, C. V., *Chem. Rev.*, **2007**, *107*, 3544-3567.
(57) Fernandez de la Mora, J., *Anal. Chim. Acta.*, **2000**, *406*, 93-104.
(58) Fischer, H.; Polikarpov, I.; Craievich, A. F., *Protein Sci.*, **2004**, *13*, 2825-2828.
(59) Wu, Z., *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, *100*, 8880-8885.
(60) Baba, M.; Imai, T.; Nishimura, M.; Kakizaki, M.; Takagi, S.; Hieshima, K.; Nomiyama, H.; Yoshie, O., *J. Biol. Chem.*, **1997**, *272*, 14893-14898.
(61) Schenauer, M. R.; Yu, Y.; Sweeney, M. D.; Leary, J. A., *J. Biol. Chem.*, **2007**, *282*, 25182-25188.
(62) Menozzi, F. D.; Pethe, K.; Bifani, P.; Soncin, F.; Brennan, M. J.; Locht, C., *Mol. Microbiol.*, **2002**, *43*, 1379-1386.

7

Conclusions

Mass spectrometry (MS) and drift time ion mobility-mass spectrometry (DT IM-MS) in combination with additional biophysical techniques have provided valuable insight into the dynamic structure and interactions of peptides and proteins. In this chapter the conclusions of this thesis are summarised.

MS and IM-MS are emerging as powerful tools in the study of proteins and peptides, enabling complexes and conformations adopted by these species to be probed in a solvent-free environment. These techniques, have lower resolution than traditional techniques such as NMR or X-ray crystallography, yet they can still provide detailed insight into dynamic systems providing 'snapshots' of unfolding and insight into interactions occurring between species. In this thesis a combination of biophysical techniques, including MS and DT IM-MS, were applied to study the structure of conformationally dynamic systems and their interactions with biomolecules.

Chapter 3 focused on the leucine zipper regions of the aggressive oncoprotein c-MYC and its binding partner MAX, considering their interaction in the absence and presence of a potential inhibitor ligand, 10058-F4. This conformationally dynamic system undergoes a disorder-to-order transition upon binding. DT IM-MS revealed the peptide heterodimers can exist in two distinct conformational families; a disordered complex and a more extended potentially 'coiled-coil' leucine zipper type complex. DT IM-MS reveals the ligand 10058-F4 preferentially inhibits the formation of the coiled-coil structure, pushing the equilibrium towards the disordered from, which is supposedly inactive. These results demonstrate the power of this technique to distinguish between and separate multiple conformations of dynamic systems, providing insight which could not have been obtained from solution based methods, such as CD spectroscopy. Studies of a shorter c-MYC peptide comprising the leucine zipper region but which does not contain the 10058-F4 drug target binding region (c-MYC-ZipΔDT) showed no significant differences in complex formation, inhibition or binding of 10058-F4, demonstrating this ligand is not as specific as initially reported. The combination of MS and DT IM-MS can provide a useful ligand screening method in the development of more specific compounds designed to disrupt the formation of the c-MYC:MAX leucine zipper.

Chapter 4 investigated the conformationally dynamic, metamorphic protein lymphotactin (Ltn) using a combination of DT IM-MS and ECD fragmentation. The results propose that the distinct monomer and dimer conformations of Ltn can be preserved and studied in the gas-phase, suggesting this combination of techniques could be highly beneficial in the study of dynamic proteins. The results provided detailed insight into the fold and unfolding of this dynamic protein, highlighting the susceptible regions of each fold and demonstrating that the intrinsically disordered (ID) tail of Ltn can associate with structural core, which is thought to confer stability to the tertiary fold. Further analysis focused on two specific mutants aimed to stabilise each fold. The findings suggested the monomer mutant (CC3) is a stable mimic

of this fold. The dimer mutant (W55D) however appears less stable as a dimer than the WT Ltn, signifying that future studies are required to design a more stable mutant for the dimeric fold. In Ltn there is evidence to suggest β-sheets may be preserved in the gas-phase, which to-date have remained elusive; Ltn therefore is an excellent candidate for future studies, such as gas-phase IR to determine if these secondary structural elements can be preserved in gaseous protein ions.

In Chapter 5 the interactions of the metamorphic chemokine lymphotactin and a model pentameric GAG, Fondaparinux (Fx) were studied. The results highlight, that despite its metamorphic nature, the functions of WT Ltn are not completely separate and defined *in vitro* and Ltn is capable of binding Fx as both a monomer and a dimer. ECD and DT IM-MS suggest the resulting complexes are adopting stable, specific conformations. Furthermore, complexes formed between Ltn and Fx appear to be stabilised by the ID tail, suggesting this preserved disordered region is functionally important with respect to GAG binding and could help explain why this region is conserved. A finding reinforced from the dramatically different behaviour of the WT 1-72 construct (which lacks the ID tail). The monomer mutant (CC3) displays similar behaviour to the WT Ltn and studies of the dimer mutant (W55D) again highlighted the necessity for a better defined mutant. The results suggest that Fx binds to a basic patch of amino acids on both the Ltn10 and Ltn40 folds. Interactions between GAGs and a basic region of a protein would enable multiple interactions to occur, strengthening and stabilising binding and the corresponding complexes formed. Studies then focused on Ltn constructs involving mutation of basic amino acids within this accessible region which were found to decrease the GAG binding propensity of Ltn, reinforcing the importance of this region for GAG binding. In future studies it would be of interest to study the interactions of Ltn with a range of GAGs of varying to chain lengths and sulfonation patterns to better understand and characterise the interactions.

Chapter 6 focused on the interactions of the chemotactic antimicrobial peptides human β-defensins (HBD) with Fx, considering HBD2 and HBD3. HBD2 solely binds Fx in a dimeric form and the findings from DT IM-MS, ECD, CID and molecular dynamics suggest these species are stable and specific. Limited aggregation is also observed upon addition of Fx to HBD2. HBD3, however, aggregates extensively in the presence of Fx, resulting in a wide range of oligomeric species being observed, with species comprising up to ten HBD3 monomers and up to eight chains of Fx detected. The binding of β-defensins with GAGs could have implications for the chemotactic activity of these peptides. This suggests that, as with the defined chemokines, interactions with cell surface GAGs may enable a

concentration gradient along which cells can migrate *in vivo* and aggregation may help overcome effects of vascular flow. As with Ltn, future studies should investigate the interactions of HBDs with a range of GAGs to further characterise these interactions. Finally, the interactions between HBD2 and HBD3 were considered and the defensins found to form high intensity, specific heterodefensin complexes, ranging from dimers to hexamers. Furthermore, dimeric, trimeric and tetrameric heterodefensin complexes were observed to bind GAGs. The conformations adopted by the heterodefensin and Fx-bound heterodefensin complexes suggest in both cases stable specific complexes are being formed. This finding is significant as the ability of β-defensins to interact with each other, and to go on to bind other biomolecules, has not to-date been reported. Further *in vitro* and *in vivo* studies would be of interest here to determine the functional relevance of the heterodefensin complexes.

The combination of MS, ECD and DT IM-MS provides a powerful approach to distinguish between multiple conformations of dynamic peptide and protein systems, and their interactions with other biomolecules, which can be difficult to distinguish in traditional solution-phase studies. The advantage of using such techniques becomes evident when considering the small sample volumes required, with as little as 25 μg of protein providing detailed insight into dynamic conformations or interactions. Furthermore, through the combination of biophysical techniques insight can be gained not only into individual proteins or peptides but also the interactions of these species and resulting complexes formed, enabling all species present to be studied simultaneously and providing insight into important biological systems.

Appendices

Appendix 1

A1.1 Amino acids

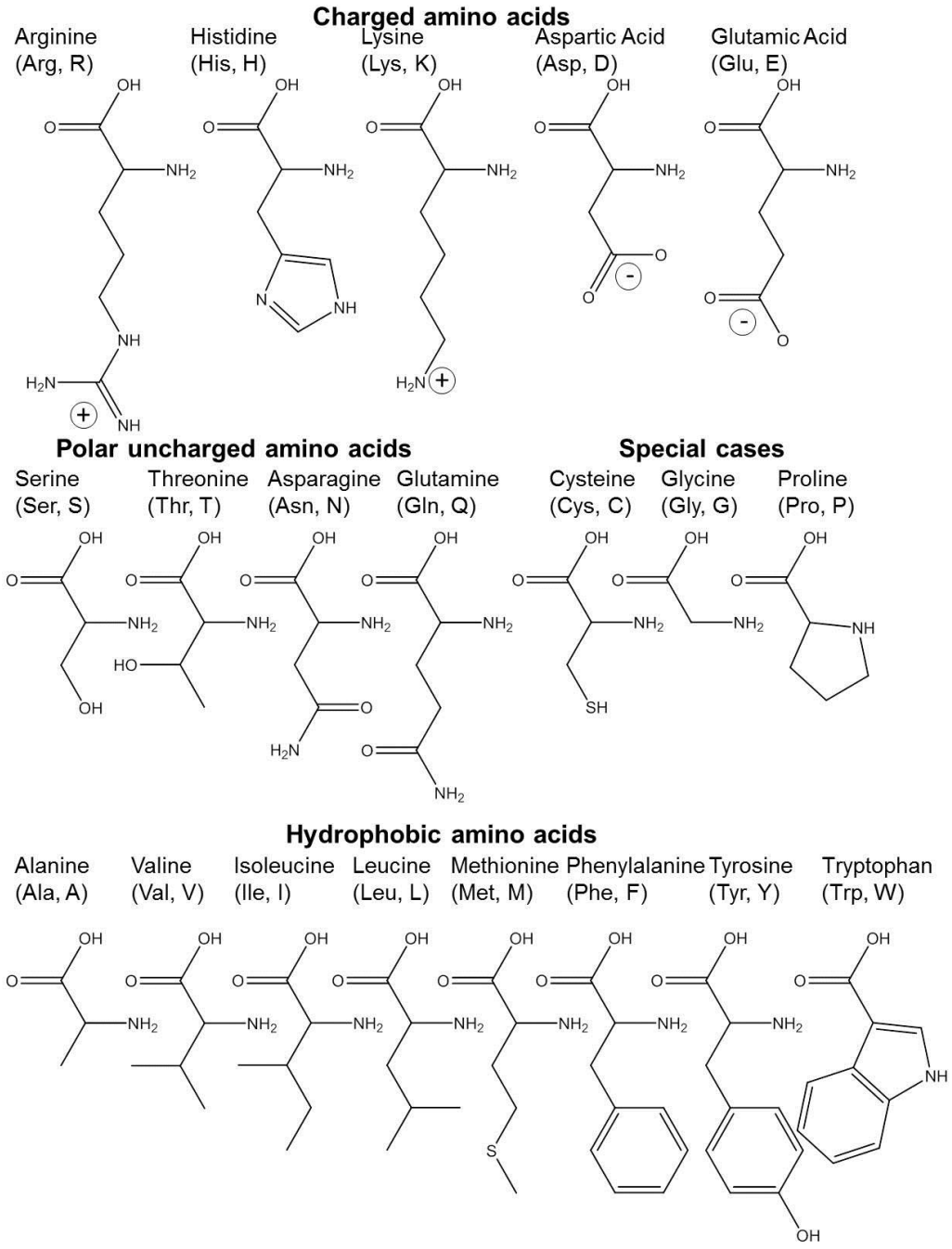

Figure A1.1: The structures of the 20 standard amino acids annotated with the three and single letter code.

A1.2 ITC analysis

In ITC for each injection the heat evolved or absorbed is proportional to the change in concentration of the ligand, see Equation A1.1, where V is the reaction volume and ΔH^0 is the enthalpy of binding.

$$q = \mathrm{V}.\Delta H^0.\Delta[L]_{bound} \quad [A1.1]$$

In a simple model where there exists a single set of identical binding sites on the molecule (M) (described by Equation A1.2) the cumulative heat (Q) can be described in terms of total ligand concentration ($[L]_{total}$), as shown in Equation A1.3. Hence, the stoichiometry of binding (n), association constant (K_a) and ΔH^0 can be determined by fitting to a graph of Q versus $[L]_{total}$, yielding a hyperbolic curve. Alternatively one can plot the 1st derivative of Q with respect to $[L]_{total}$ (incremental heat signal) against the molar ratio ($[L]_{total}/[M]_{total}$), yielding a sigmoidal curve.

$$\mathrm{M} + \mathrm{n.L} \rightleftharpoons \mathrm{ML_n} \quad [A1.2]$$

$$Q = \frac{(1+[M]_{total}.n.K_a+K_a[L]_{total})-[(1+[M]_{total}.n.K_a+K_a[L]_{total})^2-4[M]_{total}.n.{K_a}^2[L]_{total}]}{2.K_a/_{V\Delta H^0}} \quad [A1.3]$$

From Equation A1.4 the magnitudes of the free energy (ΔG^0) and entropy (ΔS^0) of binding can be determined.

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = RT\mathrm{In}K_a \quad [A1.4]$$

The change in heat capacity (ΔC_p) can also be determined if two experiments are run at difference temperatures, based on Equation A1.5.

$$\Delta \boldsymbol{C_p} = \frac{(\Delta \boldsymbol{H^0}_{T2}-\Delta \boldsymbol{H^0}_{T1})}{(\boldsymbol{T_2}-\boldsymbol{T_1})} = \frac{(\Delta \boldsymbol{S^0}_{T2}-\Delta \boldsymbol{S^0}_{T1})}{\mathbf{In}\left(\boldsymbol{T_2}/\boldsymbol{T_1}\right)} \quad [A1.5]$$

Appendix 2: c-MYC

A2.1 c-MYC peptide synthesis details

In brief, peptide carboxamides were synthesised using the Fmoc-based technique using Rink amide MBHA resin. Peptide synthesis was carried out using an ABI 433A (Applied Biosystems, Foster City, CA, USA), automated peptide synthesiser. Identity of peptides was confirmed using liquid chromatography-mass spectrometry on a Waters uPLC/SQD-LC series electrospray mass spectrometer (Waters, Manchester, UK). LC-MS was performed using a gradient of 5-95% acetonitrile containing 0.1% trifluroacetic acid (TFA) over 10 minutes (flow rate of 0.6 mL/min). Semi-preparative high performance liquid chromatography (HPLC) was performed using a Phenomenex LUNA C18 column and a gradient of 10-90% acetonitrile containing 0.1% TFA over 50 minutes (flow rate of 4.0 mL/min). c-MYC-Zip and c-MYC-ZipΔDT were prepared as acetylated peptides and MAX-Zip was non-aceteylated. Peptides were supplied as lyophilised samples and stored at -20 °C until use.

A2.2 Molecular dynamics simulations

All molecular dynamics simulations and theoretical calculations run were performed by Dr Massimilano Porrini. The coordinate file for the NMR structure of the c-MYC-ZipΔDT:MAX-Zip hetero-dimeric leucinne zipper was downloaded the Research Collaboratory for Structural Bioinformatics (RCSB) databank web site, (PDB code 1A93). Monomeric and dimeric species were built from these coordinates, by splitting the dimer into monomers and adding the four residue stretch (YILS) to create the c-MYC-Zip. The CCS of the two heterodimeric leucine zippers, c-MYC-Zip:MAX-Zip and c-MYC-ZipΔDT:MAX-Zip, were calculated from these initial coordinates files minimised *in vacuo*, using the trajectory method (TM) implemented in MOBCAL code[1].

A MD simulation was performed on the c-MYC-Zip:MAX leucine zipper complex to determine its stability at 300 K, after the addition of the additional four amino acids to this structure. The *sandar* MD engine of Amber 11 was used to propagate the trajectories, utilising the AMBER99SB-ildn force field with a time step of 2 fs. Bonds involving hydrogen atoms were constrained at their equilibrium values *via* the SHAKE algorithm[2]. After thermalizing and equilibrating the leucine zipper in a truncated octahedral box of explicit water TIP3P (with ca. 9780 water molecules and 2 Cl^- ions to neutralize the system) at T = 300 K and P = 1 atm, a 10 ns production run was carried out. After the water

molecules were stripped out, 100 sample structures were stored (1 every 100 ps) along the obtained trajectory and minimized *in vacuo*.

To derive the CCS of the disordered monomeric and dimeric species, the following simulated annealing (SA) procedure was used; initial minimisation was performed followed by dynamics for 60 ps at 800 K, gradual exponential stepwise cooling from 800 K to 0 K was then carried out with 2 ps per step followed by a final minimisation. This procedure was carried out 500 times for monomers and 1000 times for dimers. Chirality restraints were used to keep the angles at biological equilibrium values, as during the dynamics at 800 K they might incur a chirality inversion.

The molecular dynamics (MD) engine *sander* of Amber10 software package was utilized to propagate trajectories *in vacuo*, implementing the Amber99SB-ILDN force field[3]. All bonds involving hydrogen atoms were kept constrained at their equilibrium values using the SHAKE algorithm, and a time step of 1.0 fs was utilised. An "infinite" radial cut off (r_c = 999 Å) was used to evaluate the non-bonded interactions and the temperature was kept constant through the weak-coupling algorithm[4] during the dynamics propagation. In the case of the dimers a potential well was applied between the monomers centres of mass in order to prevent complete dissociation during the dynamics. The extent monomers moved was confined to a width of 107 Å, which is just larger than the distance between the centres of mass of the two monomers in their fully extended conformation and aligned in succession along their longitudinal axis.

A2.3 c-MYC CD studies

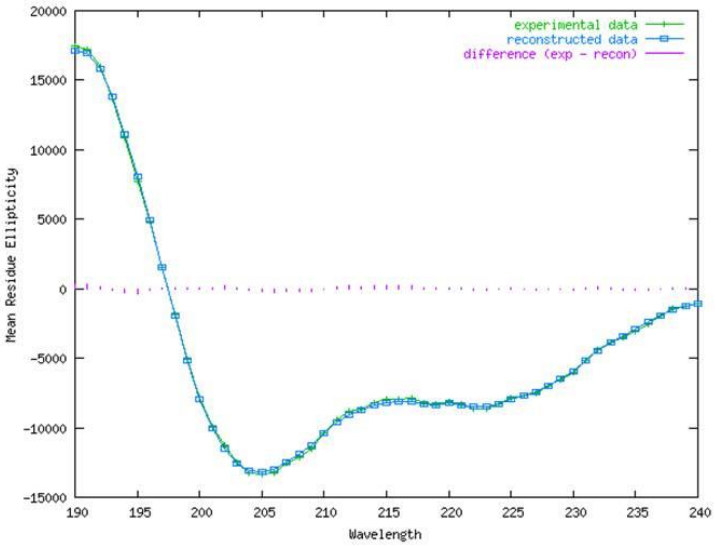

Figure A2.1: CD spectra of MAX at pH6.8, showing the fitting from the DICHROWEB server using the CDSSTR algorithm, representative for the quality of fitting in all cases.

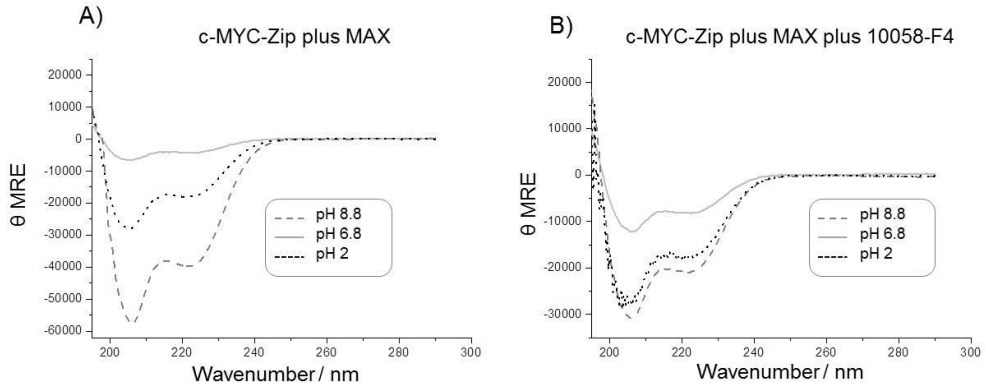

Figure A2.2: CD spectra obtained in 20 mM AmAc at pH 9.8, 6.8 and 2 of A) c-MYC-Zip plus MAX B) c-MYC-Zip plus MAX plus 10058-F4

A2.4 c-MYC mass spectrometry studies

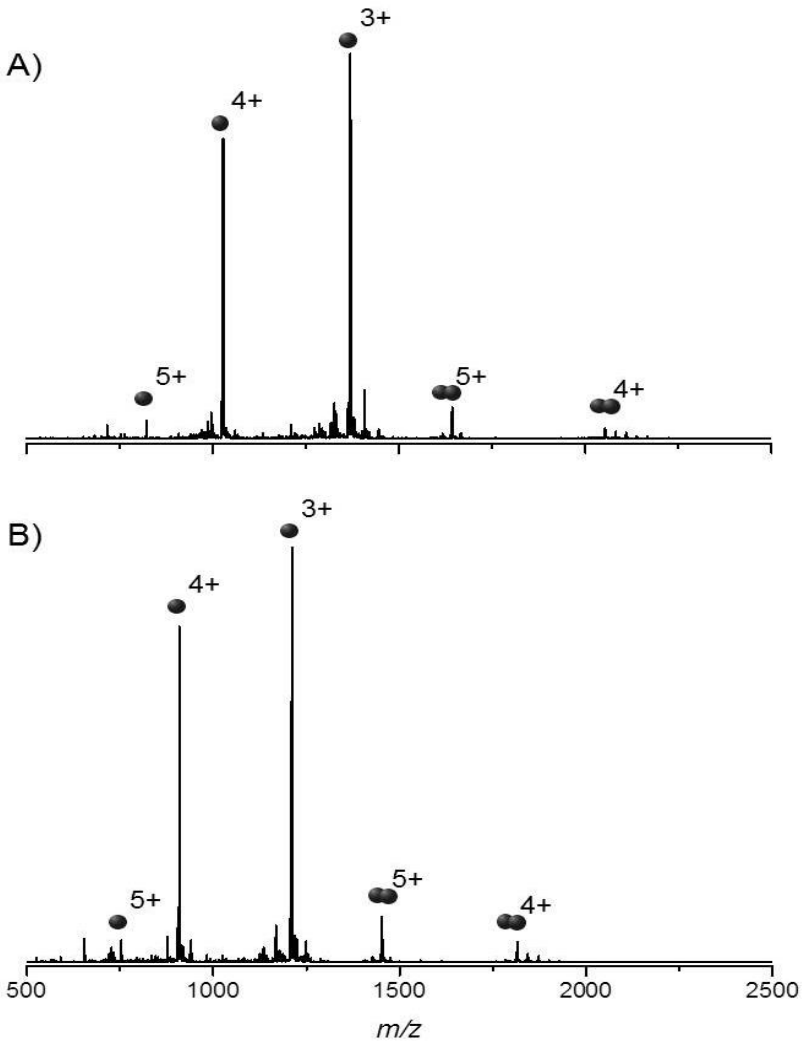

Figure A2.3: Mass spectra for the c-MYC peptides in the presence of the ligand 10058-F4 at a 1:1 ratio at 125 µM in 20 mM AmAc A) c-MYC-Zip B) c-MYC-ZipΔDT

A2.5 Attempts to observe 10058-F4 peptide complexes.

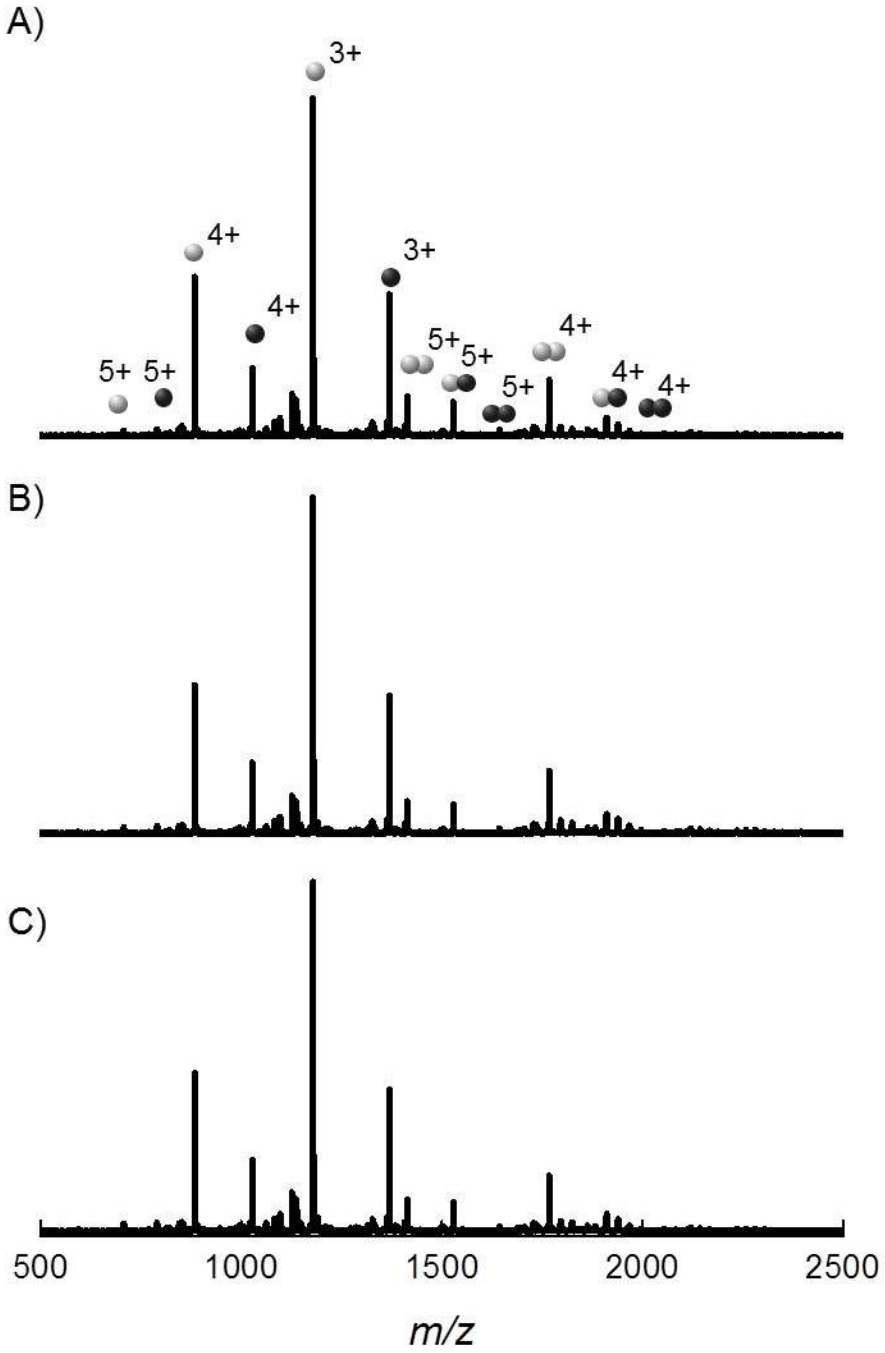

Figure A2.4: Mass spectra obtained for c-MYC plus MAX plus 10058-F4 at a 1:1:1 ratio at 125 μM concentration in 20 mM AmAc, after incubation at room temperature (18 °C) for A) 15 minutes, B) one hour and C) three hours. A sphere represents a monomeric unit of c-MYC-Zip (black) or MAX-Zip (grey).

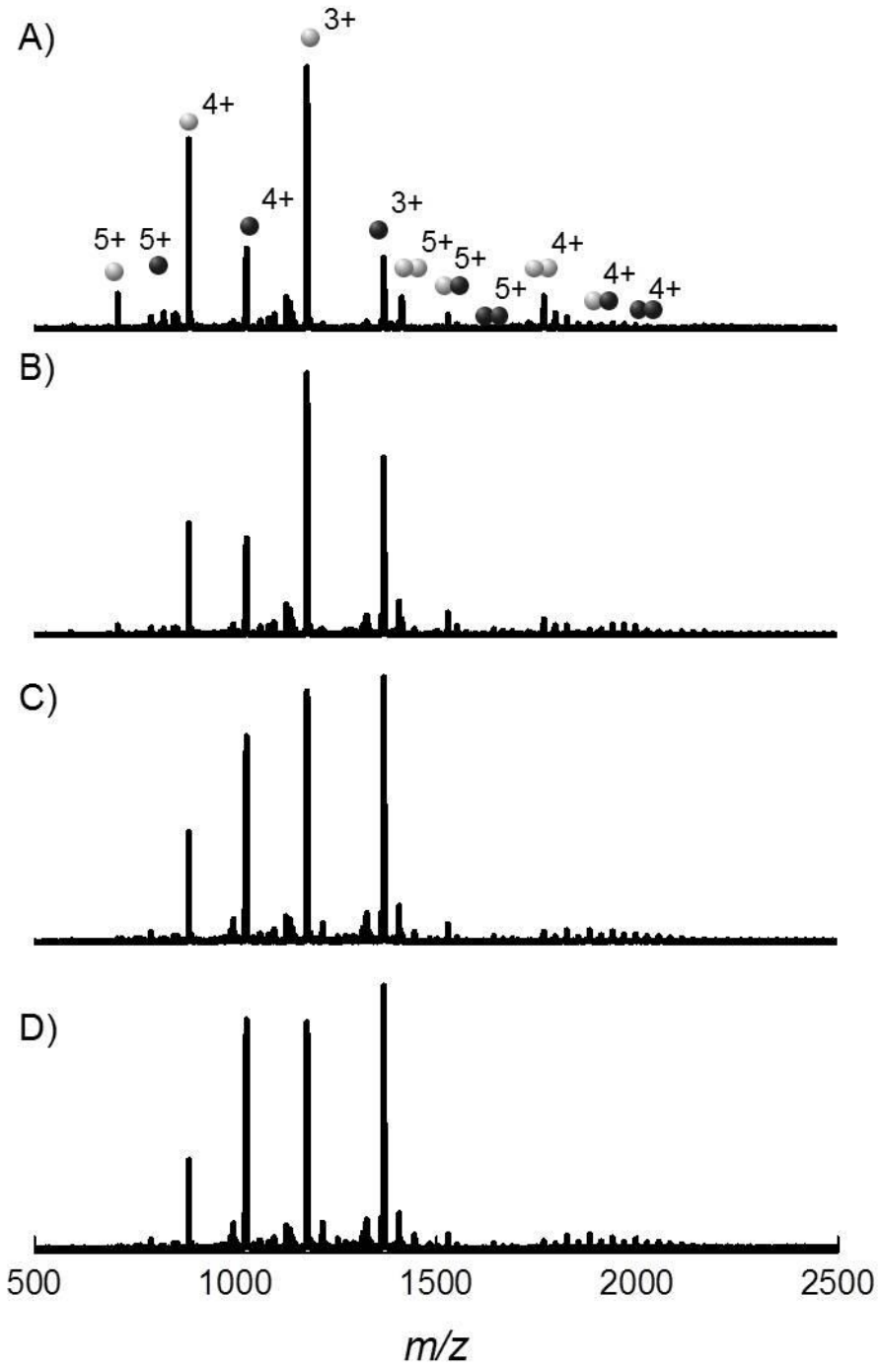

Figure A2.5: Mass spectra obtained for c-MYC plus MAX plus 10058-F4 at a 1:1:1 ratio at 125 μM concentration in A) 10 mM AmAc B) 40 mM AmAc C) 80 mM AmAc D) 100 mM AmAc. A sphere represents a monomeric unit of c-MYC-Zip (black) or MAX-Zip (grey).

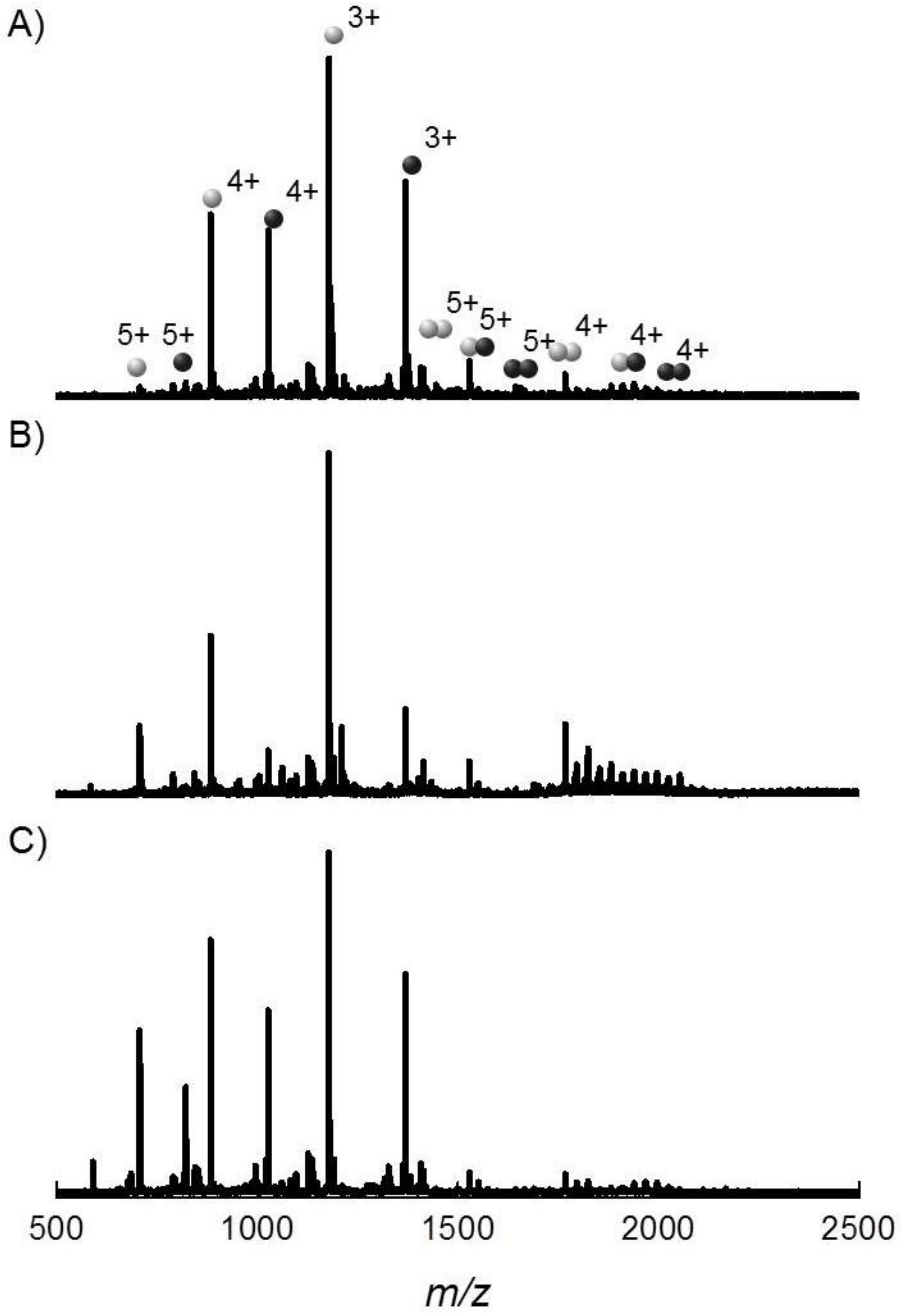

Figure A2.6: Mass spectra obtained for c-MYC plus MAX plus 10058-F4 at a 1:1:1 ratio at 125 μM concentration in 20 mM AmAc at A) pH 8.8 B) pH 5.8 and C) pH 2.8 . pH was altered through the addition of NaOH (pH 8.8) or formic acid (pH 5.8 and 2.8). A sphere represents a monomeric unit of c-MYC-Zip (black) or MAX-Zip (grey).

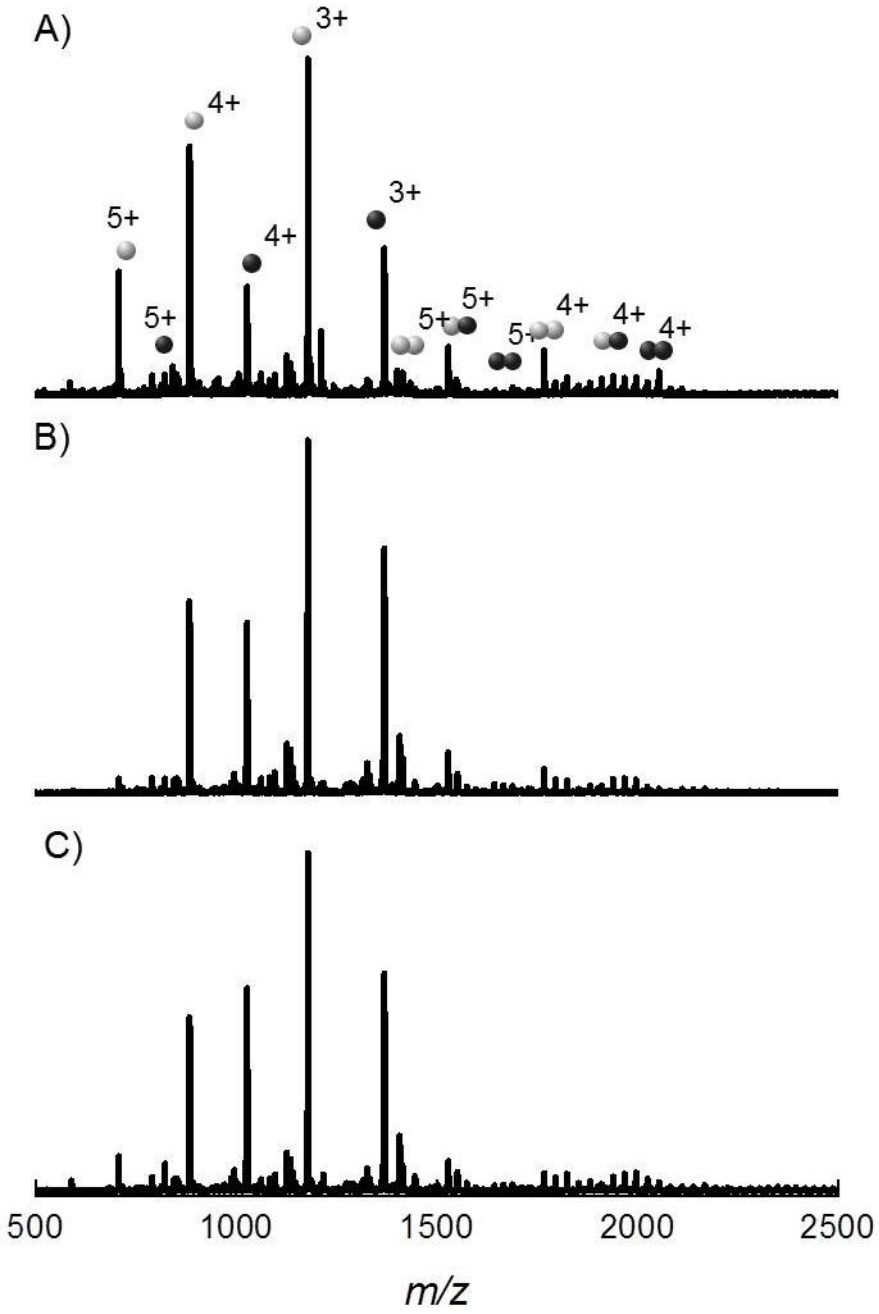

Figure A2.7: Mass spectra obtained for c-MYC plus MAX plus 10058-F4 at a 1:1:1 ratio at 125 μM concentration in 20 mM AmAc. Spectra acquired with a capillary voltage of A) 1.3 kV B) 1.7 kV and C) 2.0 kV. A sphere represents a monomeric unit of c-MYC-Zip (black) or MAX-Zip (grey).

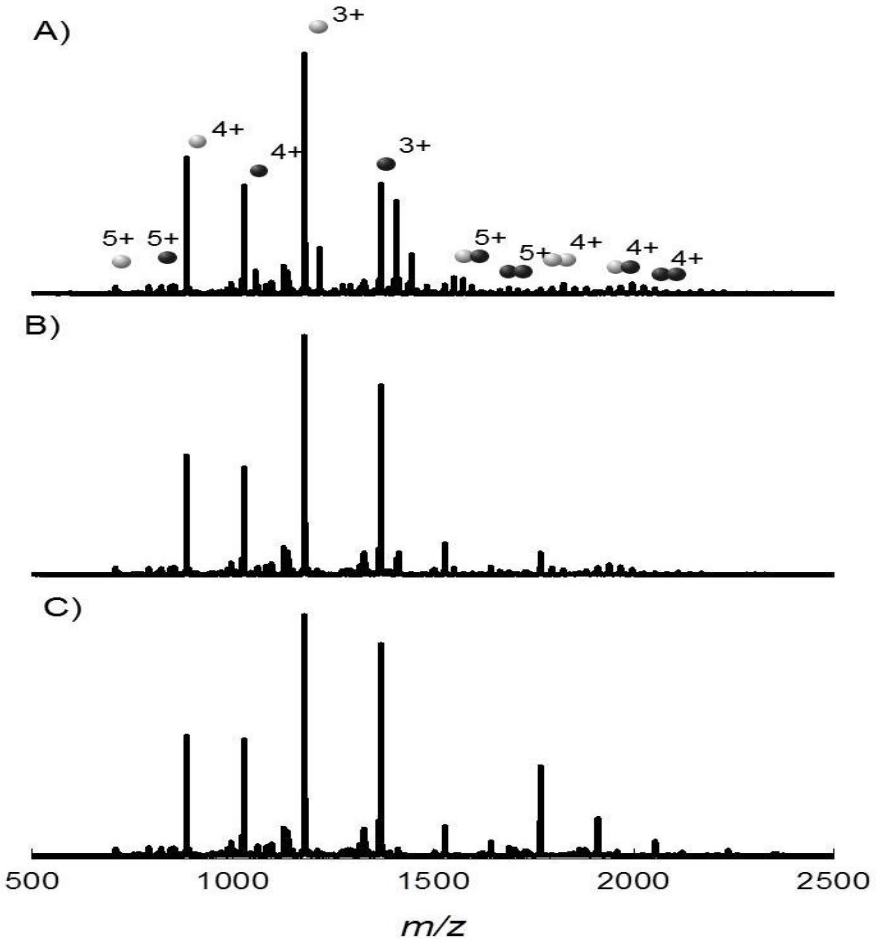

Figure A2.8: Mass spectra obtained for c-MYC plus MAX plus 10058-F4 at a 1:1:1 ratio at 125 μM concentration in 20 mM AmAc. Spectra acquired with a cone voltage of A) 15 V B) 25 V and C) 40 V. A sphere represents a monomeric unit of c-MYC-Zip (black) or MAX-Zip (grey).

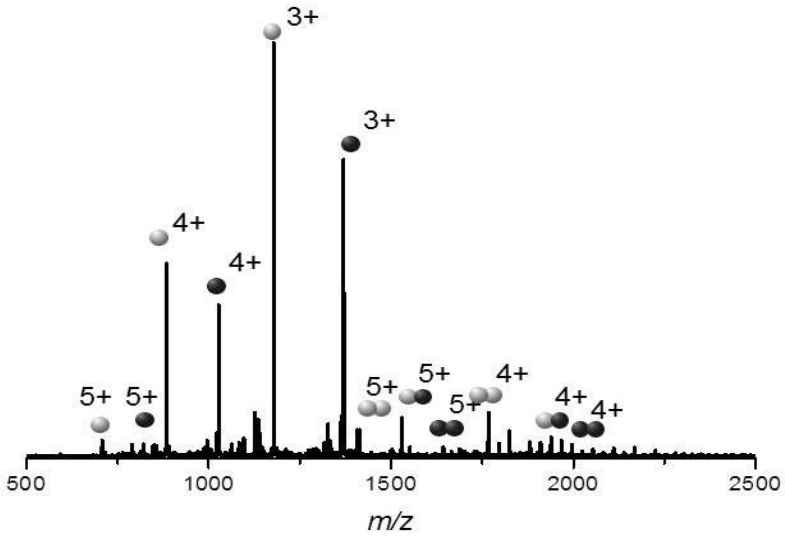

Figure A2.9: Mass spectra obtained for c-MYC plus MAX plus 10058-F4 at a 1:1:1 ratio at 125 μM concentration in 20 mM AmAc. Spectra acquired during solvent vapour exposure with acetonitrile. A sphere represents a monomeric unit of c-MYC-Zip (black) or MAX-Zip (grey).

A2.6 CCS of c-MYC, MAX homo- and heterodimers in presence of 10058-F4

Species		*Experimental CCS, Å² (absence of ligand)*	*Experimental CCS, Å² (presence of ligand)*
Max-Zip	$[M+4H]^{4+}$	565 ± 2	564 ± 8
	$[M+3H]^{3+}$	554 ± 6	546 ± 2
	$[D+4H]^{4+}$	857 ± 12	866 ± 5
c-Myc-ZipΔDT	$[M+4H]^{4+}$	629 ± 4	632 ± 8
	$[M+3H]^{3+}$	595 ± 5	586 ± 4
	$[M+2H]^{2+}$	453 ± 11	443 ± 8
	$[D+5H]^{5+}$	941 ± 8	950 ± 2
c-Myc-Zip	$[M+4H]^{4+}$	711 ± 2	716 ± 9
	$[M+3H]^{3+}$	649 ± 3	633 ± 4
	$[M+2H]^{2+}$	479 ± 8	n/a
	$[D+5H]^{5+}$	1131 ± 5	1122 ± 8

Figure A2.1: Experimental CCS for c-MYC-Zip, c-MYC-ZipΔDT and MAX-Zip monomers (M) and homodimers (D), determined in the absence and presence of the ligand 10058-F4.

Appendix 3: Lymphotactin equilibrium studies

A3.1 Molecular dynamics simulations

All molecular dynamics simulations were performed by Dr Massimilano Porrini and are described in brief below. Two independent MD runs in explicit solvent were conducted for the monomer Ltn10 (PDB 1J9O), using different initial velocities and the same starting structure coordinates. The simulation protocol implemented was as follows; the monomer was immersed in a truncated octahedron box, with minimum distance between one edge of the box and one atom of the solute of ~15 Å, the system was minimised, gradually heated up to 300 K, with density equilibrated at 300 K and 1 bar, a further 10 ns equilibration was run after which the 150 ns production run was carried out. The force field implemented was Amber99SB-ILDN[3] together with the model TIP3P[5] for water molecules, the radial cut-off for the non-bonded interactions was 8 Å, a time step of 2.0 fs was utilised, with all the bonds involving hydrogen atoms kept at their equilibrium distance, and the trajectories were propagated with the module *pmemd.cuda* of Amber12[6]. During the heating and density equilibration the solute was kept restrained with a force constant of 0.5 kcal/mol/Å^2.

The net charge of Ltn10 at pH=7 is +9 therefore in order to study the species $[M_{WT}+7H]^{7+}$ two basic residues have to be deprotonated, determined by calculating the $pK_{1/2}$ values of all residues. In order to do this one structure was taken every 10 ns for 50 ns from the first simulation in explicit solvent, these were then submitted to the H++ webserver to calculate the $pK_{1/2}$ values, the average for the five structures was then taken and showed that the lowest $pK_{1/2}$ are those of Lys41 and Lys65 and hence these amino acids were deprotonated. Water molecules and counter-ions were stripped out from the last snapshot of the second simulation in water of Ltn10 and the two protons removed. The resulting $[M+7H]^{7+}$ was firstly minimised, then gradually heated up to 300 K after which a constant temperature production run of 150 ns was carried out *in vacuo*. Along the heating and production dynamics a time step of 1.0 fs was implemented and an "infinite" (*rcut*=999 Å) radial cut-off was utilised for the non-bonded interactions.

Next Ltn40 (PDB 2JP1) was considered, as with Ltn10 two independent trajectories in explicit solvent were first run. The PDB file contains 20 models as NMR candidate structures of the dimer in water. The first and the second model were utilised as starting coordinates for the two runs, applying the same simulation protocol as that implemented for the monomer, except production runs were 200 ns long. The net charge of the dimer Ltn40 at pH=7 is +10, in order to study the 11+ species one needs to add a proton to an acidic residue.

To select which acidic residue is most likely to be protonated, the $pK_{1/2}$ values of each single acidic and basic residue belonging to the dimer were determined from 5 snapshots from the first 50 ns of the first simulation in water. This procedure shows that the weakest acidic moieties are the Glu3 of the chain A and hence this residue was protonated to form the $[M+11H]^{11+}$ species for further study. Water molecules and counter-ions were stripped out from the last snapshot of the first simulation in water of Ltn40, following which the proton was added to the above glutamic acid. The resulting $[D+11H]^{11+}$ was subjected to the same procedure of gas-phase MD as the monomer $[M+7H]^{7+}$.

In all cases salt bridge occupancies were derived using the software *cpptraj*[7] of AmberTools13[6]. The command utilised was *hbond* imposing a radial cut-off between hydrogen bond donor and acceptor equal to 3 Å. The computation considered only the anionic carboxylate of acidic residues (Asp and Glu) and the cationic ammonium of basic residues (Lys and Arg). Occupancies of greater than 20 % were reported in each case.

A3.2 WT Ltn solution studies

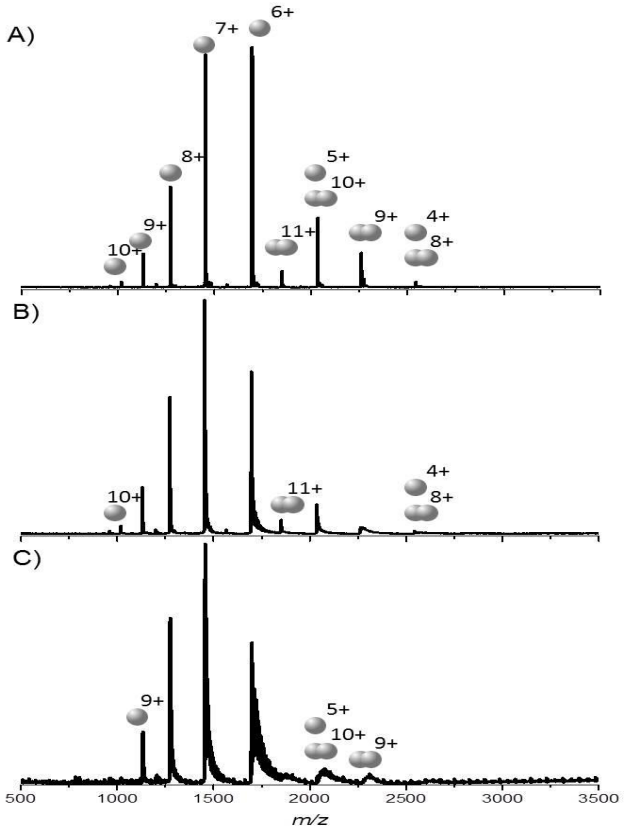

Figure A3.1: Mass spectra obtained for 50 μM Ltn in 20 mM AmAc with A) 0 mM NaCl B) 1 mM NaCl C) 5 mM NaCl. For top panels all major species are labelled, in all other panels the highest and lowest charge states observed for monomer and dimer are labelled.

A3.3 Experimental CCS for full length WT Ltn and WT Ltn 1-72

Species	Experimental CCS /Å²
$[M_{WT}+4H]^{4+}$	757 ± 4
$[M_{WT}+5H]^{5+}$	967 ± 13
$[M_{WT}+6H]^{6+}$	1021 ± 27
$[M_{WT}+7H]^{7+}$	1100 ± 63
$[M_{WT}+8H]^{8+}$	1528 ± 8
$[M_{WT}+9H]^{9+}$	1560 ± 40
$[M_{WT}+9H]^{9+*}$	1685

Table A3.1: The experimental CCS for all observed species of monomeric WT Ltn. *represents species observed in all repeats but only clearly resolvable in one. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these.

Species	Experimental CCS /Å²
$[M_{1-72}+4H]^{4+}$	856 ± 4
$[M_{1-72}+5H]^{5+}$	916 ± 8
$[M_{1-72}+6H]^{6+}$	950 ± 9
$[M_{1-72}+7H]^{7+}$ *	1037
$[M_{1-72}+7H]^{7+}$	1229 ± 12
$[M_{1-72}+8H]^{8+}$	1416 ± 8
$[D_{1-72}+8H]^{8+}$	1404 ± 6
$[D_{1-72}+9H]^{9+}$	1425 ± 8
$[D_{1-72}+10H]^{10+}$	1603
$[D_{1-72}+11H]^{11+}$	1751 ± 10
$[D_{1-72}+11H]^{11+}$	1872 ± 4
$[D_{1-72}+12H]^{12+}$	1861 ± 8

Table A3.2: The experimental CCS for all observed species of monomeric WT 1-72 Ltn. *represents species observed in all repeats but only clearly resolvable in one. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these.

Species	Experimental CCS /Å²
$[D_{WT}+8H]^{8+}$	1376 ± 25
$[D_{WT}+9H]^{9+}$	1605 ± 12
$[D_{WT}+9H]^{9+*}$	1534
$[D_{WT}+10H]^{10+}$	1778 ± 47
$[D_{WT}+11H]^{11+*}$	1822
$[D_{WT}+12H]^{12+*}$	1894
$[D_{WT}+13H]^{13+*}$	1902

Table A3.3: The experimental CCS for all observed species of dimeric WT Ltn. * represents species observed in all repeats but only clearly resolvable in one. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these.

A3.4 Ability of the structural core of Ltn (WT 1-72) to bind the ID tail (WT 72-93), as shown by mass spectrometry

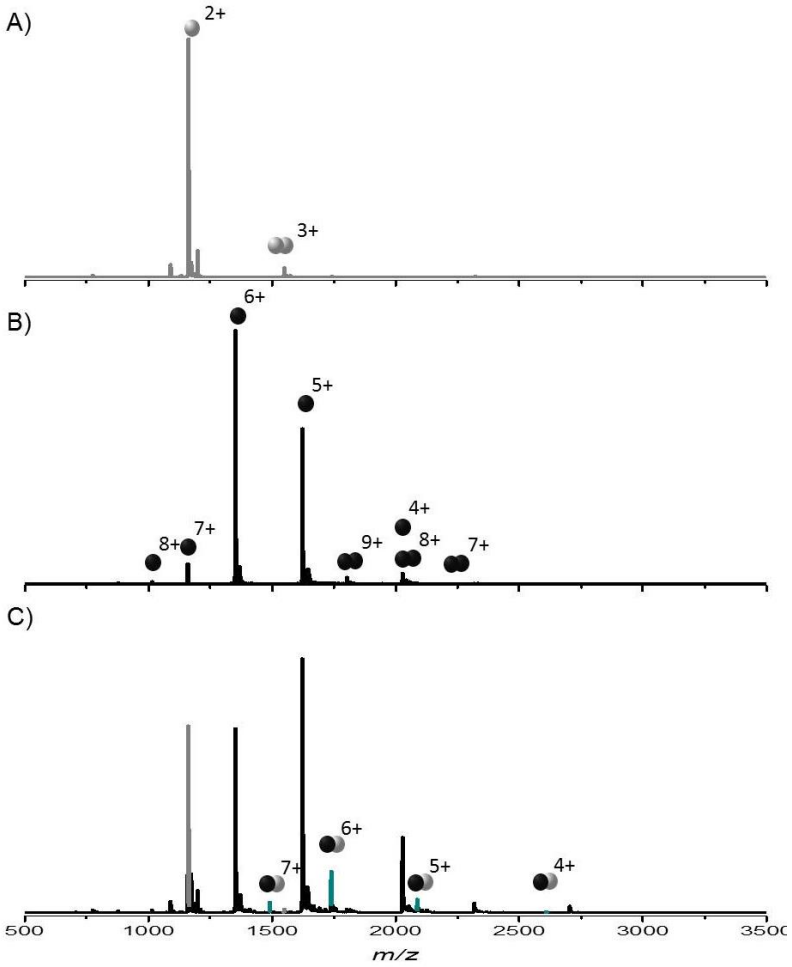

Figure A3.2: Mass spectra obtained at 50 μM in 20 mM AmAc for A) WT 72-93 B) WT 1-72 and C) 1:1 mixture of WT 72-93 and WT 1-72 both at 50 μM. In all cases peaks corresponding to WT 72-93 are shown in grey, WT 1-72 is shown in black and complex peaks are shown in cyan.

A3.5 MS spectra for WT Ltn acquired on the FT ICR-MS instrument used for ECD studies

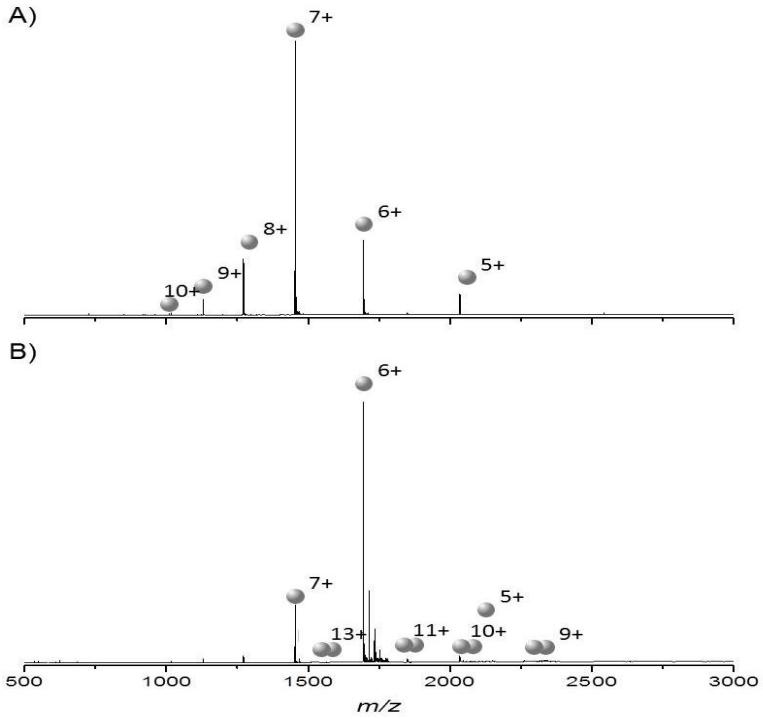

Figure A3.3: Mass spectra obtained on a 12T Apex Qe FTICR for WT Ltn A) Acquired from electrospray ionisation of 10 μM Ltn in 100 mM ammonium acetate B) Acquired from nano-electrospray ionisation of 30 μM Ltn in 100 mM ammonium acetate

A3.6 Ltn10 salt bridge studies

Residue 1	Residue 2	Run 1 occupancy (%)	Run 2 occupancy (%)
Glu30(OE2)	Arg56(NH2)	52.15	
Glu30(OE1)	Arg56(NH1)	48.53	
Glu30(OE1)	Arg56(NH2)	45.17	
Asp6(OD2)	Arg8(NH2)	41.52	
Glu30(OE2)	Arg56(NH1)	36.65	
Asp6(OD2)	Arg8(NE)	33.99	
Asp6(OD1)	Arg8(NH2)	33.66	
Asp6(OD1)	Arg8(NE)	31.26	
Asp57(OD2)	Arg17(NH2)	26.75	32.94
Asp6(OD1)	Arg17(NH2)	23.65	
Asp6(OD2)	Arg17(NH1)	21.14	
Glu3(OE2)	Arg8(NH2)		48.91
Glu3(OE1)	Arg8(NH2)		45.02
Glu3(OE1)	Arg8(NE)		35.08
Glu3(OE2)	Arg8(NE)		32.05
Asp57(OD1)	Arg17(NH2)		28.83
Asp57(OD1)	Arg60(NH2)		26.40
Asp57(OD2)	Arg17(NH1)		24.93
Asp57(OD1)	Arg17(NH1)		21.07

Table A3.4: Ltn10 salt bridge occupancies ≥ 20 % along the first and second 150 ns MD trajectory in explicit solvent

Residue 1	Residue 2	Occupancy (%)
Asp63(OD2)	Arg64(NE)	95.22
Glu3(OE1)	Arg8(NE)	732.21
Asp57(OD1)	Arg60(NH2)	66.24
Glu3(OE2)	Arg8(NH2)	61.87
Asp57(OD2)	Arg17(NH2)	59.46
Asp57(OD2)	Arg17(NH1)	55.95
Asp57(OD1)	Arg60(NE)	53.35
Asp63(OD1)	Arg64(NH2)	52.67
Asp57(OD2)	Arg60(NH2)	51.85
Asp63(OD2)	Arg64(NH2)	49.76
Asp57(OD1)	Arg17(NH1)	37.73
Asp57(OD1)	Arg17(NH2)	37.26
Asp57(OD2)	Arg60(NE)	35.17
Glu3(OE1)	Arg8(NH2)	34.59

Table A3.5: Ltn10 $[M_{WT}+7H]^{7+}$ salt bridge occupancies ≥ 20 % along the 150 ns MD trajectory *in vacuo*

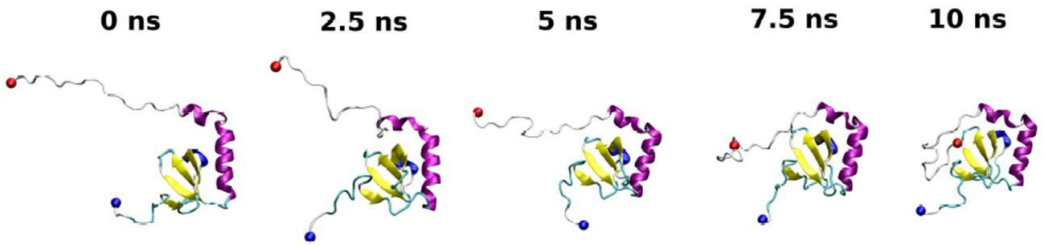

Figure A3.4: Snapshots of the structure of Ltn10 taken every 2.5 ns over the 10 ns equilibrium period during the second simulation in water.

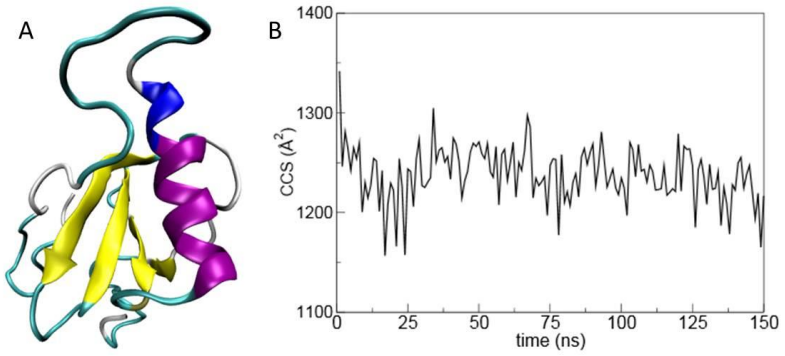

Figure A3.5: A) Ltn10 $[M + 7H]^{7+}$ at the end of the *in vacuo* simulation B) CCS of $[M + 7H]^{7+}$with respect to time evolution along the 150 ns MD *in vacuo*.

Visual inspection of the PDB file suggests the possibility of formation of an additional classical salt bridge between Glu30-Arg34 and Arg34-Asp49 and MD simulations in solvent support this possibility (Figure A3.6A) however this is present at lower occupancy (< 10%) than those listed in Tables A3.4 and A3.5. However this salt bridge can strengthen in the gas-phase (Figure A3.6B), determined by selecting two structures one at 60 ns and one at 70 ns (top panel Figure A3.6A) and subjecting them to a 100 ns *in vacuo* production run.

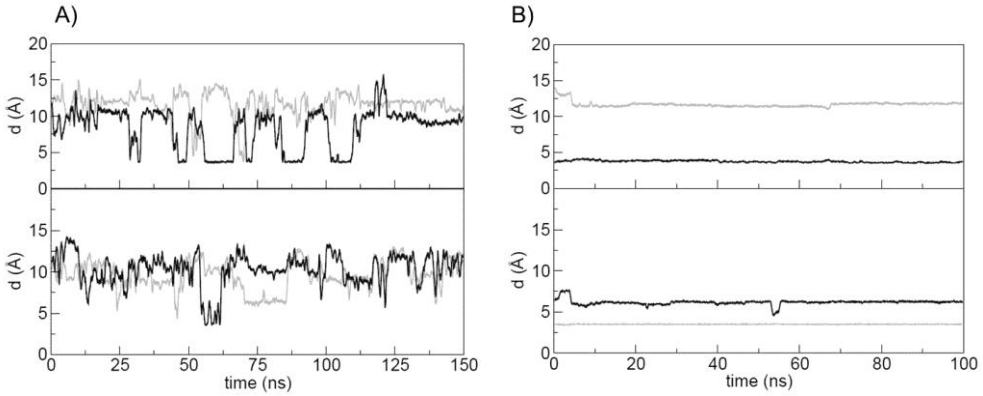

Figure A3.6: Distance between the centres of mass of the negative moiety of Glu30 side chain and positive moiety of Arg34 side chain (grey line) and between the centres of mass of the positive moiety of Arg34 side chain and negative moiety of Asp49 side chain (black line), as a function of time. A) For the first (bottom) and second (top) simulation in water. B) *in vacuo* simulations where, the starting structure was taken at 60 ns (top graph) and at 70 ns (bottom graph) of the second simulation of Ltn10 in explicit water

A3.7 WT Ltn $[M+10H]^{10+}$ trapping time experiments.

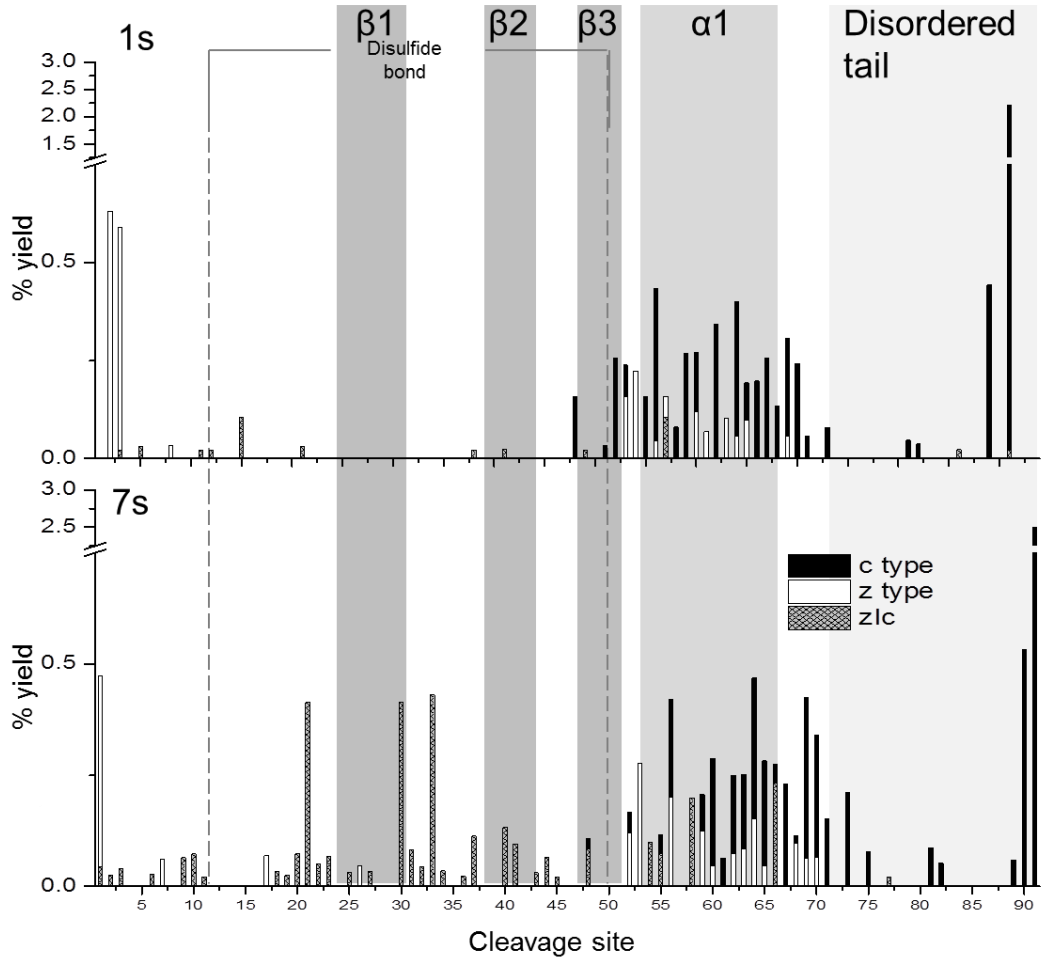

Figure A3.7: Percentage yields calculated for zIc, c and z type fragments as a function of cleavage site for $[M_{wt}+10H]^{10+}$ after one (top) and seven (bottom) seconds trapping prior to ECD.

The decreased trapping time for $[M_{WT}+10H]^{10+}$ was enabled using different instrumental parameters than those used to obtain the fragmentation maps in Chapter 4 Section 4.3.3, and a concentration six times higher therefore differences in fragmentation maps could also be due to differences in instrumental settings, other than trapping times.

A3.8 Ltn40 salt bridge studies

Residue 1	Residue 2	Run 1 Occupancy (%)	Run 2 Occupancy (%)
Intra-chain A			
Asp49(OD1)	Arg8(NE)	44.77	
Asp49(OD2)	Arg8(NH2)	37.09	
Asp49(OD1)	Arg8(NH2)	29.57	
Asp49(OD2)	Arg8(NE)	27.12	
Asp49(OD2)	Arg34(NH2)	23.87	42.60
Asp49(OD1)	Arg34(NH2)		39.55
Asp49(OD2)	Arg34(NE)		34.71
Asp49(OD2)	Lys7(NZ)		27.58
Asp49(OD1)	Arg34(NE)		27.19
Intra-chain B			
Asp49(OD2)	Arg34(NH2)	58.34	25.25
Asp49(OD1)	Arg8(NE)	51.83	22.99
Asp49(OD2)	Arg34(NE)	34.84	22.00
Asp49(OD1)	Arg8(NH2)	31.48	20.99
Asp49(OD2)	Arg8(NH2)	28.05	
Asp49(OD1)	Arg34(NH2)		26.77
Asp49(OD1)	Arg34(NE)		25.67
Asp49(OD2)	Arg34(NE)		23.60
Asp49(OD2)	Arg8(NE)		23.09
Asp57(OD1)	Arg56(NH2)		22.21
Asp57(OD1)	Arg56(NE)		21.12
Asp57(OD2)	Arg56(NE)		20.92
Asp6(OD1)	Lys7(NZ)		20.84
Inter-chain A,B			
Glu30(OE1)	Arg22(NH2)	37.39	24.28
Arg(NH1)	Glu30(OE2)	33.15	
Glu30(OE2)	Arg17(NH2)	32.51	76.75
Glu30(OE2)	Arg17(NH1)	32.31	31.40
Arg22(NH2)	Glu30(OE1)	32.05	
Arg22(NH2)	Glu30(OE2)	29.24	
Glu30(OE2)	Arg22(NE)	23.46	
Glu30(OE1)	Arg22(NE)	23.18	54.11
Glu30(OE2)	Arg22(NH2)	21.78	56.46
Arg22(NH1)	Glu30(OE1)	21.68	
Lys24(NZ)	Glu30(OE2)		42.60
Lys24(NZ)	Glu30(OE1)		38.28
Glu30(OE1)	Lys24(NZ)		24.35

Table A3.6: Ltn40 salt bridge occupancies ≥ 20 % along the first and second 200 ns MD trajectory in explicit solvent

Residue 1	Residue 2	Occupancy (%)
Intra-chain A		
Asp49(OD2)	Arg8(NH2)	49.48
Asp49(OD2)	Arg8(NE)	48.80
Asp49(OD1)	Arg34(NH2)	48.62
Asp49(OD1)	Arg8(NE)	48.32
Asp49(OD2)	Arg34(NH2)	48.28
Asp49(OD1)	Arg8(NH2)	45.81
Asp49(OD1)	Arg34(NE)	32.61
Asp49(OD2)	Arg34(NE)	30.81
Intra-chain B		
Asp49(OD2)	Arg34(NH2)	97.52
Asp49(OD1)	Arg8(NH2)	96.67
Asp57(OD2)	Arg56(NH2)	96.65
Asp49(OD1)	Arg8(NE)	95.11
Asp6(OD2)	Lys7(NZ)	93.90
Asp49(OD2)	Arg34(NE)	87.22
Asp57(OD2)	Arg56(NE)	82.91
Inter-chain A,B		
Arg22(NH2)	Glu30(OE1)	97.35
Arg22(NH1)	Glu30(OE2)	94.50
Glu30(OE1)	Lys24(NZ)	92.28
Arg22(NE)	Asp57(OD1)	91.41
Arg22(NH2)	Asp57(OD1)	89.83
Asp57(OD2)	Arg17(NH1)	86.65
Arg42(NH2)	Glu3(OE1)	86.91
Arg17(NH2)	Glu3(OE1)	85.19
Glu30(OE2)	Arg61(NH2)	84.16
Glu30(OE2)	Arg61(NH1)	83.63
Glu30(OE1)	Arg22(NE)	61.25
Glu30(OE2)	Arg22(NE)	50.72
Arg17(NE)	Glu3(OE2)	36.59
Glu30(OE2)	Arg22(NH2)	33.46
Arg17(NE)	Glu3(OE1)	24.75

Table A3.7: Ltn40 salt bridge occupancies ≥ 20 % along the 200 ns MD trajectory *in vacuo*

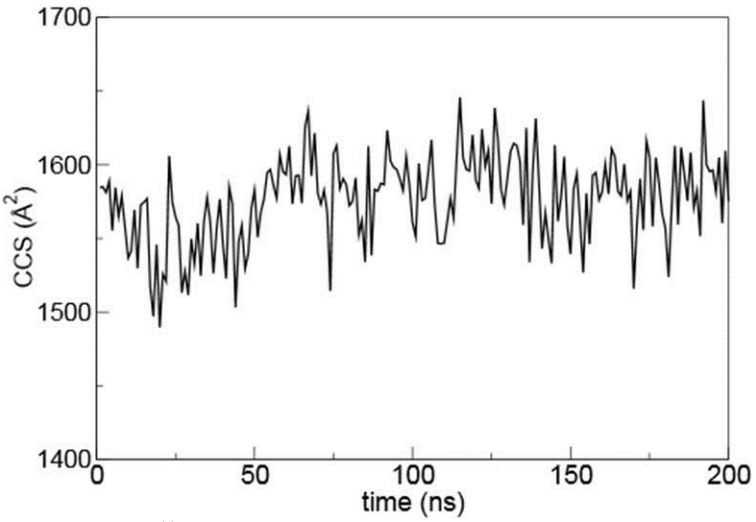

Figure A3.8: CCS of $[D + 11H]^{11+}$ with respect to time evolution along the 200 ns MD *in vacuo*

A3.9 Experimental CCS for CC3

Species	*Experimental CCS /Å²*	*'Supercharged' experimental CCS /Å²*
$[M_{CC3}+5H]^{5+}$	1012 ± 2	1023 ± 8
$[M_{CC3}+6H]^{6+}$	1014 ± 3	1046 ± 15
$[M_{CC3}+7H]^{7+}$	1163 ± 4	1252 ± 30
$[M_{CC3}+8H]^{8+}$	1243 ± 5	1439 ± 25
$[M_{CC3}+9H]^{9+}$		1498 ± 10
$[M_{CC3}+10H]^{10+}$		1622 ± 17
$[M_{CC3}+11H]^{11+}$		1697 ± 1
$[M_{CC3}+12H]^{12+}$		1757 ± 28
$[M_{CC3}+13H]^{13+}$		1824 ± 6
$[D_{CC3}+9H]^{9+}$	1611 ± 2	
$[D_{CC3}+10H]^{10+}$	1802 ± 5	

Table A3.8: The experimental CCS for all observed species of monomeric CC3, obtained from buffered and 'supercharged' solution conditions. The average values obtained from three different day repeats are reported here and errors are reported as the standard deviation between these.

A3.10: W55D solution studies

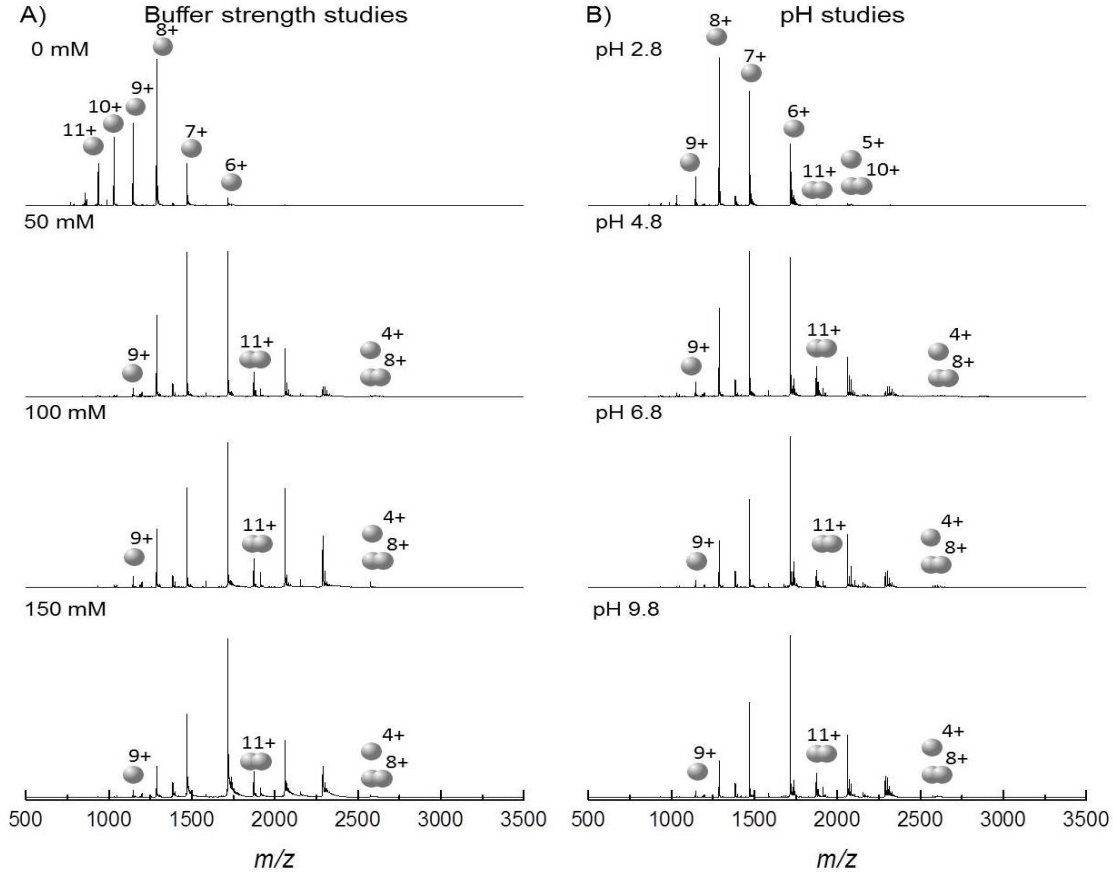

Figure A3.9: W55D Ltn mass spectra for 50 μM Ltn. A) buffer strength from 0-150 mM AmAc, B) 20 mM AmAc at pH 2.8-9.8, altered through addition of formic acid or ammonia. For top panels all major species are labelled, in all other panels the highest and lowest charge states for monomer and dimer are labelled.

A3.11 Experimental CCS for W55D

Species	*Experimental CCS /Å²*
$[M_{W55D}+4H]^{4+}$	778 ± 8
$[M_{W55D}+5H]^{5+}$	839 ± 7
$[M_{W55D}+6H]^{6+}$	1057 ±12
$[M_{W55D}+7H]^{7+}$	1150 ± 26
$[M_{W55D}+8H]^{8+}$	1401 ± 6
$[M_{W55D}+9H]^{9+}$	1407 ± 9
$[D_{W55D}+8H]^{8+}$	1509 ± 6
$[D_{W55D}+9H]^{9+}$	1641 ± 8
$[D_{W55D}+9H]^{9+*}$	1674
$[D_{W55D}+10H]^{10+}$	1704 ± 6
$[D_{W55D}+11H]^{11+}$	1829 ± 19
$[D_{W55D}+11H]^{11+*}$	2020
$[D_{W55D}+12H]^{12+}$	2062 ± 8
$[D_{W55D}+13H]^{13+}$	2103 ± 67
$[D_{W55D}+13H]^{13+}$	2237 ± 95
$[D_{W55D}+14H]^{14+}$	2246 ± 5
$[D_{W55D}+15H]^{15+*}$	2469
$[D_{W55D}+16H]^{16+}$	2519 ± 12

Table A3.9: The experimental CCS for all observed species of W55D. * represents species observed in all repeats but only clearly resolvable in one. The average values obtained from three repeats are reported here and errors are reported as the standard deviation between these.

Appendix 4: Ltn:GAG binding studies

A4.1 Images of n-ESI capillaries

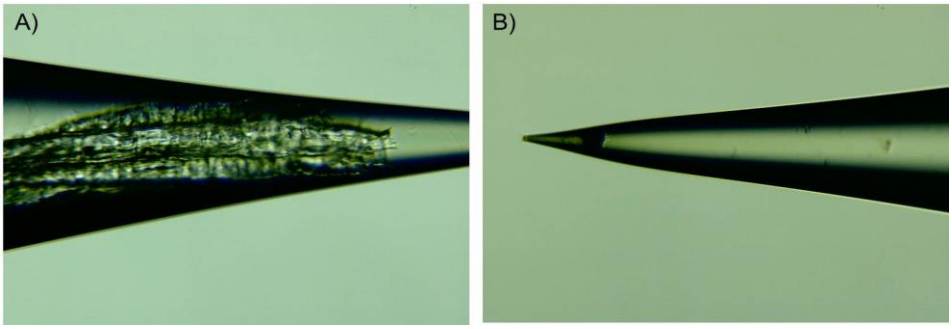

Figure A4.1: Images of n-ESI of 1:1 solutions of Ltn plus Fx at 50 μM concentration in 20 mM AmAc. A) W55D and B) WT 1-72 Ltn.

A4.2 Studies into WT Ltn:GAG binding

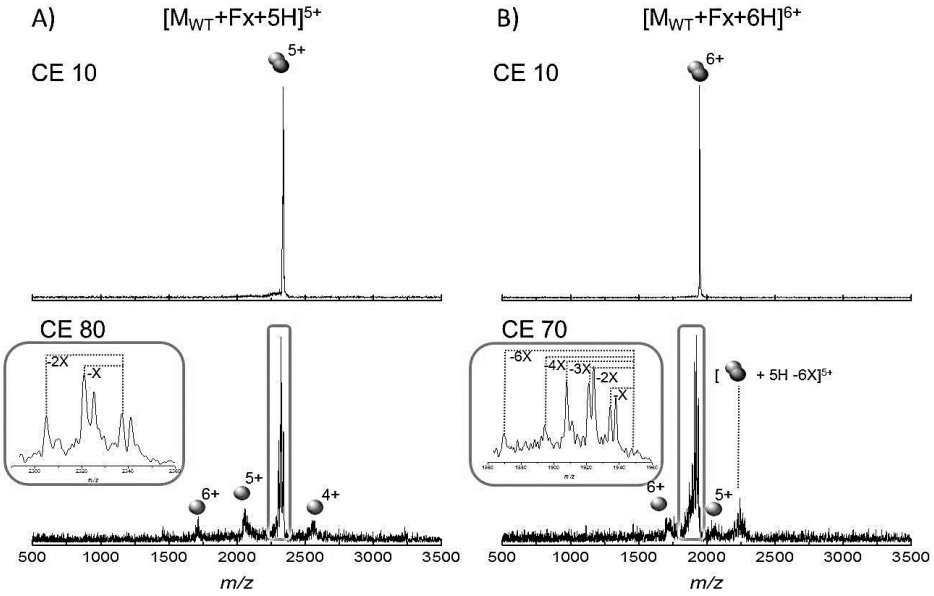

Figure A4.2: Spectra obtained for isolated species (top) and after CID (bottom) for A) $[M_{WT}+Fx+5H]^{5+}$ and B) $[M_{WT}+Fx+6H]^{6+}$. Acquired from a 1:1 (Ltn:Fx) solution prepared at 50 μM concentration in 20 mM AmAc. Where X represents (SO_3).

Species	Expected mass (Da)	Observed mass (Da)
WT Ltn Monomer (M_{WT})	10172.6	10171.4
WT Ltn Dimer (D_{WT})	20345.2	20344.6
WT Ltn Monomer + Fx (M_{WT}+Fx)	11680.8	11680.2
WT Ltn Dimer + Fx (D_{WT}+Fx)	21853.4	21851.9
WT Ltn Dimer +2 Fx (D_{WT}+2Fx)	23361.6	23360.7

Table A4.1: Expected and observed masses for all WT Ltn and WT Ltn+Fx complexes, averaged over all observed charge states in the spectra in Chapter 5 Figure 5.2B. The expected mass of fully desalted and protonated Fx is 1508.19 Da.

Species	*Experimental CCS /Å²*
$[M_{WT}+Fx+5H]^{5+}$	845 ± 27
$[M_{WT}+Fx+6H]^{6+}$	1004 ± 8
$[M_{WT}+Fx+7H]^{7+}$	1247 ± 54
$[D_{WT}+Fx+9H]^{9+}$	1530 ± 1
$[D_{WT}+Fx+10H]^{10+}$	1602 ± 7
$[D_{WT}+2Fx+8H]^{8+}$	1426 ± 31
$[D_{WT}+2Fx+9H]^{9+}$	1591 ± 37
$[D_{WT}+2Fx+10H]^{10+*}$	1651

Table A4.2: The experimental CCS for all observed Fx bound species of WT Ltn. * represents species observed in all repeats but only resolvable in one. The average values obtained from three different day repeats are reported here and errors are reported as the standard deviation between these.

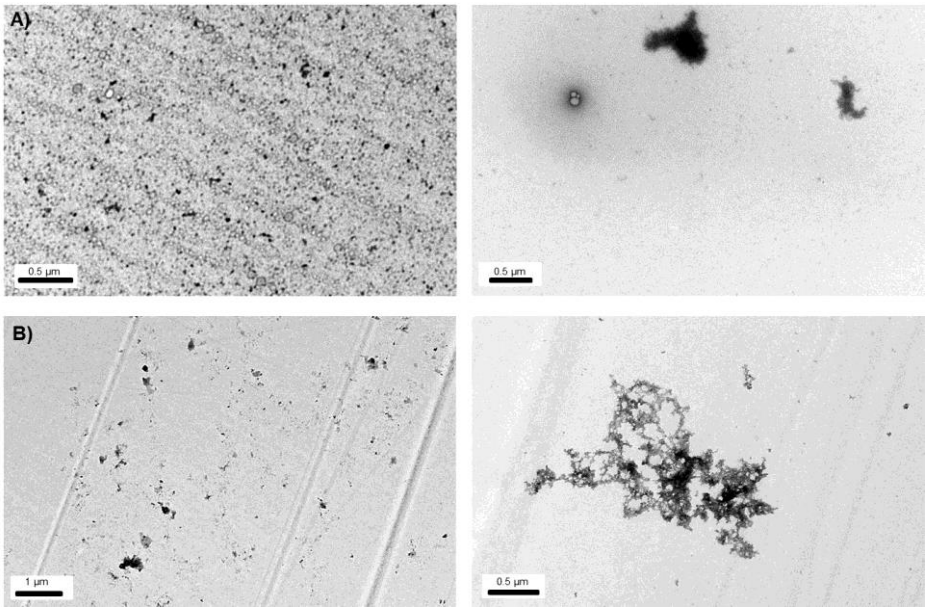

Figure A4.3: Representative TEM images obtained for A) WT Ltn and B) Foundaparinux. Each prepared at 50 µM concentration in 20 mM AmAc.

A4.3 Studies into WT 1-72 Ltn:GAG binding

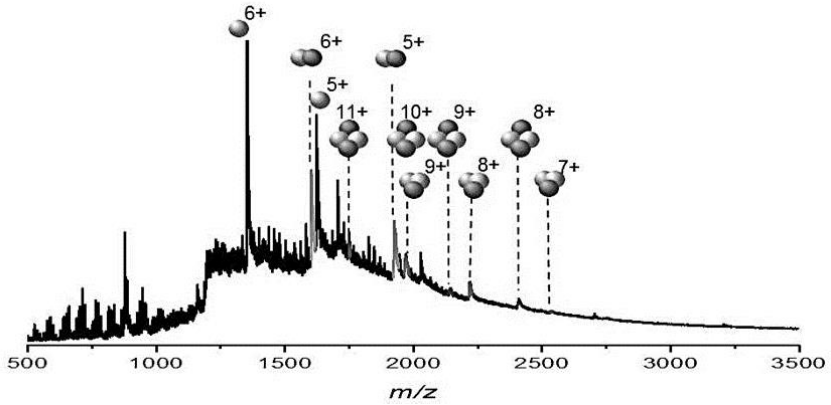

Figure A4.4: Spectra obtained for a 1:1 mixture of WT 1-72 plus Fx, at 50 µM protein concentration in 20 mM AmAc.

Species	Expected mass (Da)	Observed mass (Da)
WT 1-72 Ltn Monomer (M_{1-72})	8111.3	8111.6
WT 1-72 Ltn Dimer (D_{1-72})	16222.6	16223.9
WT 1-72 Ltn Monomer + Fx (M_{1-72}+Fx)	9619.5	19618.2
WT 1-72 Ltn Dimer + Fx (D_{1-72}+Fx)	17730.8	17730.4
WT 1-72 Ltn Dimer +2 Fx (D_{1-72}+2Fx)	19239.0	19239.21

Table A4.3: Expected and observed masses for all WT 1-72 Ltn and WT 1-72 Ltn+Fx complexes, averaged over all observed charge states in the spectra in Chapter 5 Figure 5.6B. The expected mass of fully desalted and protonated Fx is 1508.19 Da.

Species	*Experimental CCS /Å²*
$[M_{1-72}+Fx+4H]^{4+*}$	946
$[M_{1-72}+Fx+5H]^{5+}$	995 ± 16
$[M_{1-72}+Fx+6H]^{6+}$	1232 ± 46
$[D_{1-72}+Fx+7H]^{7+}$	1403 ± 25
$[D_{1-72}+Fx+7H]^{7+}$	1660 ± 12
$[D_{1-72}+Fx+8H]^{8+}$	1420 ± 19
$[D_{1-72}+Fx+8H]^{8+}$	1788 ± 5
$[D_{1-72}+Fx+9H]^{9+}$	1624 ± 20
$[D_{1-72}+Fx+9H]^{9+}$	1884 ± 48
$[D_{1-72}+Fx+9H]^{9+}$	2153 ± 24

Table A4.4: The experimental CCS for all observed Fx bound species of WT 1-72 Ltn. * represents species observed in all repeats but only resolvable in one. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these.

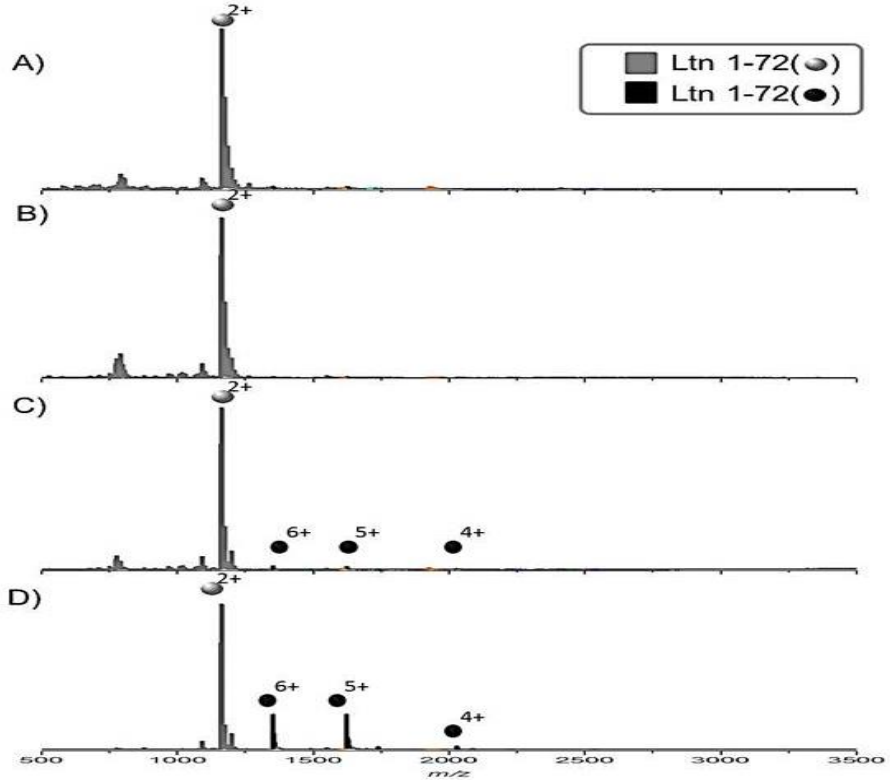

Figure A4.5: Spectra obtained for WT 1-72 plus WT 72-93 plus Fx in 20 mM AmAc A) prepared at a 1:1:1 ratio at 50 µM concentration B) prepared at a 2:2:1 (1-72:72-93:Fx) ratio at 50 µM WT 1-72 concentration C) prepared at a 2:2:1 (1-72:72-93:Fx) ratio at 25 µM WT 1-72 concentration D) prepared at a 4:4:1 (1-72:72-93:Fx) ratio at 25 µM WT 1-72 concentration

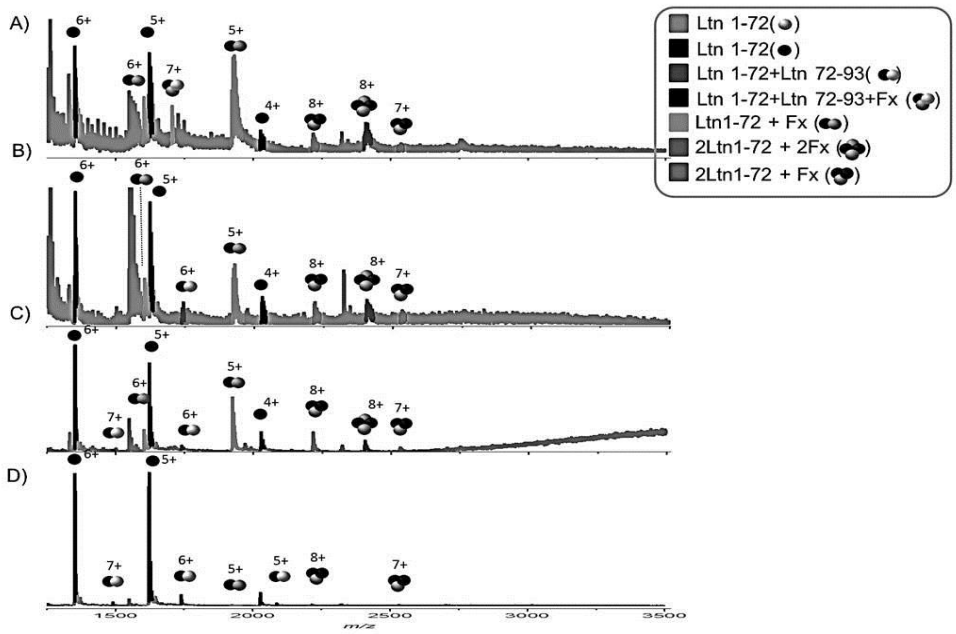

Figure A4.6: Zoom in of the 1250-3500 m/z region of the spectra obtained for WT 1-72 plus WT 72-93 plus Fx in 20 mM AmAc, as shown in Figure A3.6 A) prepared at a 1:1:1 ratio at 50 µM concentration B) prepared at a 2:2:1 (1-72:72-93:Fx) ratio at 50 µM WT 1-72 concentration C) prepared at a 2:2:1 (1-72:72-93:Fx) ratio at 25 µM WT 1-72 concentration D) prepared at a 4:4:1 (1-72:72-93:Fx) ratio at 25 µM WT 1-72 concentration.

A4.4 Studies into CC3:GAG binding

Species	Expected mass (Da)	Observed mass (Da)
CC3 Monomer (M_{CC3})	10183.6	10184.1
CC3 Dimer (D_{CC3})	20367.2	23068.0
CC3 Monomer + Fx (M_{CC3}+Fx)	11691.8	11692.7
CC3 Dimer + Fx (D_{CC3}+Fx)	21875.4	21876.3
CC3 Dimer +2 Fx (D_{CC3}+2Fx)	23383.6	23384.7

Table A4.5: Expected and observed masses for CC3 and CC3+Fx complexes, averaged over all observed charge states in the spectra in Chapter 5 Figure 5.10B. The expected mass of fully desalted and protonated Fx is 1508.19 Da.

Species	*Experimental CCS /Å²*
$[M_{CC3}+Fx+5H]^{5+}$	833 ± 7
$[M_{CC3}+Fx+6H]^{6+}$	1075 ± 21
$[M_{CC3}+Fx+7H]^{7+}$	1105 ± 13
$[D_{CC3}+Fx+8H]^{8+}$	1424 ± 28
$[D_{CC3}+Fx+9H]^{9+}$	1524 ± 16
$[D_{CC3}+Fx+10H]^{10+}$	1716 ± 4
$[D_{CC3}+2Fx+8H]^{8+}$	1423
$[D_{CC3}+2Fx+9H]^{9+}$	1624 ± 13
$[D_{CC3}+2Fx+10H]^{10+}$	1712 ± 5

Table A4.6: The experimental CCS for all observed Fx bound species of CC3. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these.

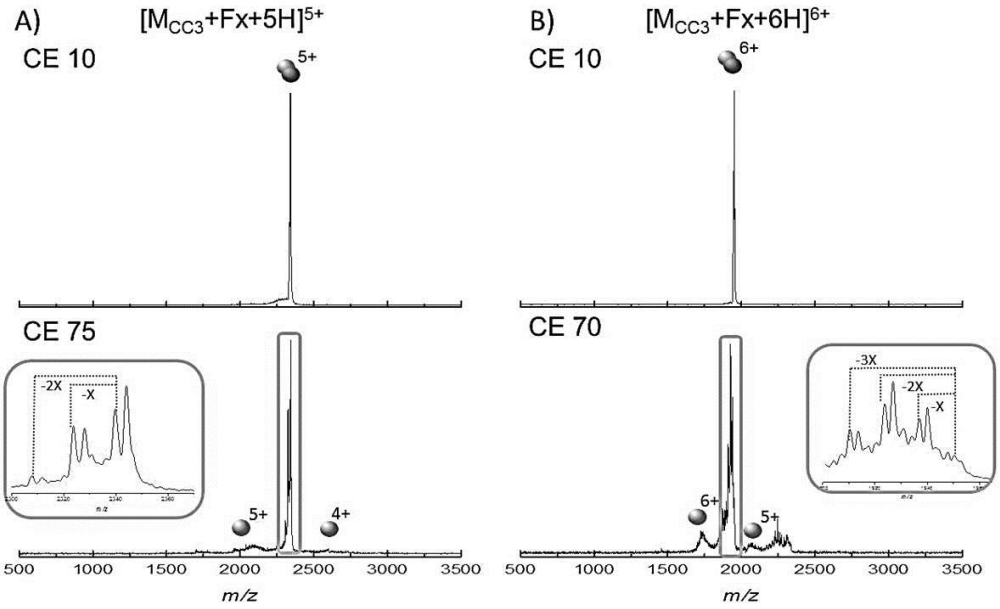

Figure A4.7: Spectra obtained for isolated species (top) and after CID (bottom) for A) $[M_{CC3}+Fx+5H]^{5+}$ and B) $[M_{CC3}+Fx+6H]^{6+}$. Acquired from a 1:1 (Ltn:Fx) solution prepared at 50 μM protein concentration in 20 mM AmAc. Where X represents (SO_3).

A4.5 Studies into W55D:GAG binding

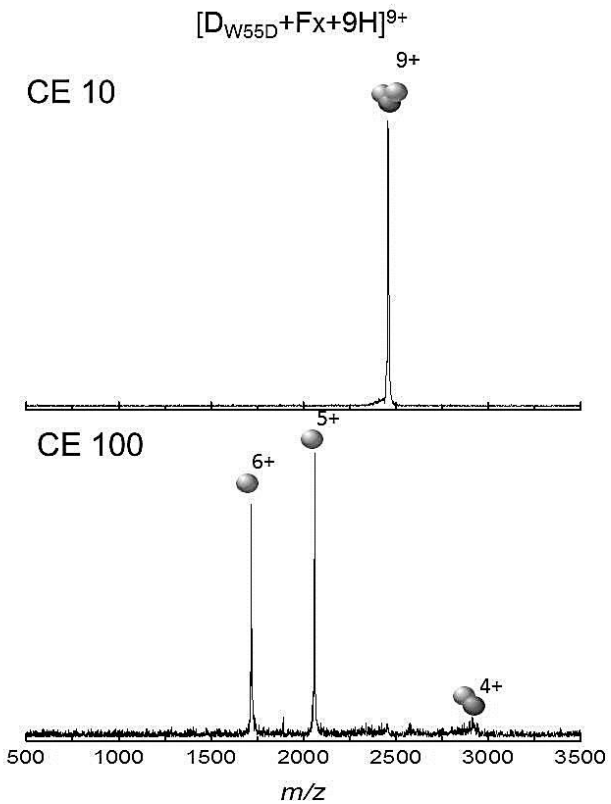

Figure A4.8: Spectra obtained for isolated species (top) and after CID (bottom) for $[D_{W55D}+Fx+9H]^{9+}$. Acquired from a 1:1 (Ltn:Fx) solution prepared at 50 μM protein concentration in 20 mM AmAc. Where X represents (SO_3).

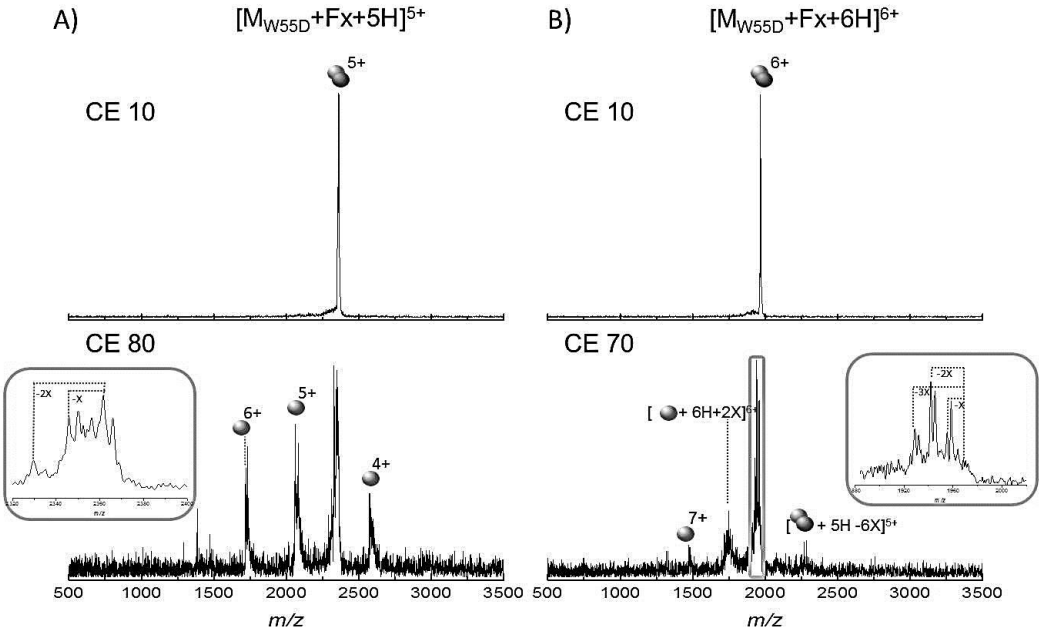

Figure A4.9: Spectra obtained for isolated species (top) and after CID (bottom) for A) [M$_{W55D}$+Fx+5H]$^{5+}$ and B) [M$_{W55D}$+Fx+6H]$^{6+}$. Acquired from a 1:1 (Ltn:Fx) solution prepared at 50 μM protein concentration in 20 mM AmAc. Where X represents (SO_3).

Species	Expected mass (Da)	Observed mass (Da)
W55D Monomer (M$_{W55D}$)	10294.7	10295.1
W55D Ltn Dimer (D$_{W55D}$)	20589.4	20590.3
W55D Ltn Monomer + Fx (M$_{W55D}$+Fx)	11802.9	11803.5
W55D Ltn Dimer + Fx (D$_{W55D}$+Fx)	22097.6	22098.9
W55D Ltn Dimer +2 Fx (D$_{W55D}$+2Fx)	23605.8	23606.6

Table A4.7: Expected and observed masses for all W55D and W55D+Fx complexes, averaged over all observed charge states in the spectra in Chapter 5 Figure 5.13B. The expected mass of fully desalted and protonated Fx is 1508.19 Da.

Species	***Experimental CCS /Å^2***
[M$_{W55D}$+Fx+5H]$^{5+}$	1162 ± 22
[M$_{W55D}$+Fx+6H]$^{6+}$	1173 ± 37
[M$_{W55D}$+Fx+7H]$^{7+}$	1253 ± 31
[D$_{W55D}$+Fx+9H]$^{9+}$	1651 ± 35
[D$_{W55D}$+Fx+10H]$^{10+}$	1874 ± 26
[D$_{W55D}$+Fx+11H]$^{11+}$	1982 ± 16
[D$_{W55D}$+Fx+12H]$^{12+*}$	2010
[D$_{W55D}$+2Fx+9H]$^{9+*}$	1760
[D$_{W55D}$+2Fx+10H]$^{10+}$	1778 ± 39
[D$_{W55D}$+2Fx+11H]$^{11+}$	1904 ± 9
[D$_{W55D}$+2Fx+12H]$^{12+*}$	2168

Table A4.8: The experimental CCS for all observed Fx bound species of W55D. * represents species observed in all repeats but only resolvable in one. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these.

A4.6 Studies into arginine Ltn mutants

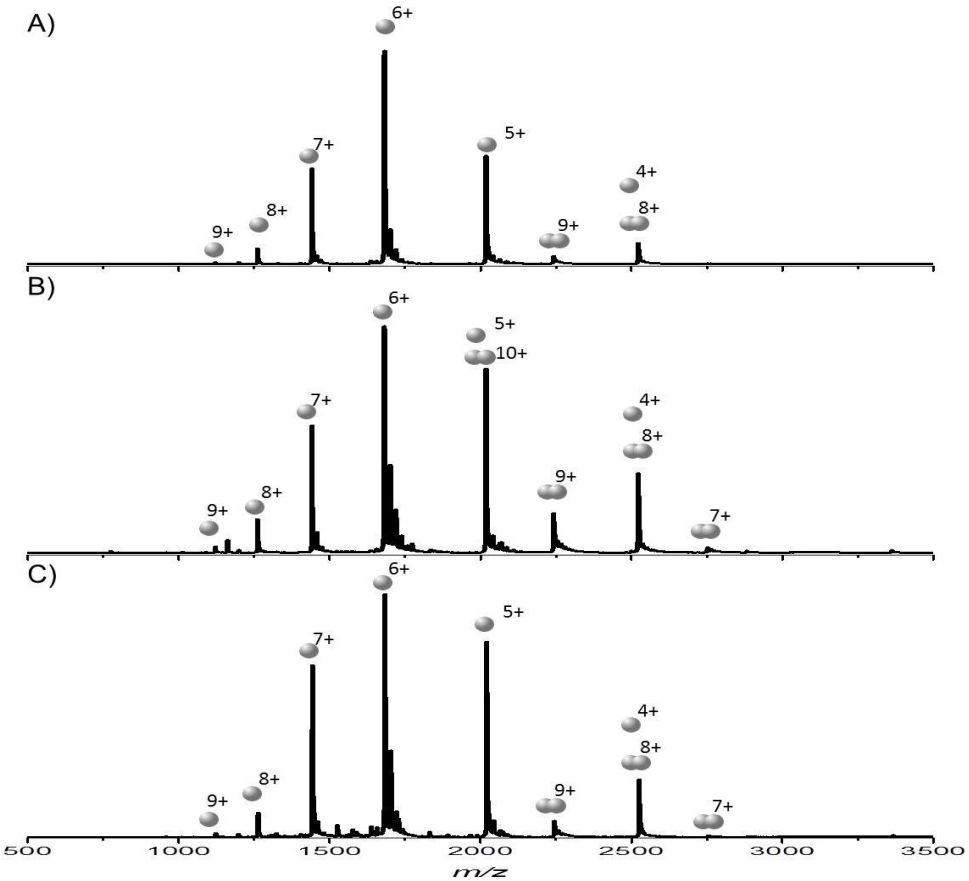

Figure A4.10: Spectra obtained for A) R23A , B) R43A and C) R23A/R43A. In all cases spectra were obtained from 50 μM protein in 20 mM AmAc.

Species	Expected mass (Da)	Observed mass (Da)
R23A Monomer (M_{R23A})	10085.6	10085.4
R23A Dimer (D_{R23A})	20171.2	20170.8
R23A Monomer+Fx (M_{R23A}+Fx)	11593.8	11593.0
R23A Dimer+Fx (D_{R23A}+Fx)	21679.4	21679.6
R23A Dimer+2Fx(D_{R23A}+2Fx)	23187.6	23187.6
R43A Monomer (M_{R43A})	10085.6	10085.7
R43A Dimer (D_{R43A})	20171.2	20170.2
R43A Monomer+Fx (M_{R43A}+Fx)	11593.8	11594.6
R43A Dimer+Fx (D_{R43A}+Fx)	21679.4	21680.4
R43A Dimer+2Fx(D_{R43A}+2Fx)	23187.6	23189.0
R23A/R43A Monomer ($M_{R23A/R43A}$)	10099.5	10099.5
R23A/R43A Dimer ($D_{R23A/R43A}$)	20199.0	20200.4
R23A/R43A Monomer+Fx ($M_{R23A/R43A}$+Fx)	11607.19	11608.3
R23A/R43A Dimer+Fx ($D_{R23A/R43A}$+Fx)	21707.2	21707.9
R23A/R43A Dimer+2Fx ($D_{R23A/R43A}$+2Fx)	23215.4	23216.5
R23A/R43A Trimer+Fx ($T_{R23A/R43A}$+Fx)	31806.7	31805.8
R23A/R43A Trimer+2Fx ($T_{R23A/R43A}$+2Fx)	33314.9	33315.7
R23A/R43A Tetramer+2Fx($TE_{R23A/R43A}$+2Fx)	43414.4	43414.9
R23A/R43A Tetramer+3Fx($TE_{R23A/R43A}$+3Fx)	44922.6	44924.5

Table A4.9: Expected and observed masses for all arginine mutant complexes, averaged over all observed charge states in the spectra in Chapter 5 Figure 5.17. The expected mass of fully desalted and protonated Fx is 1508.19 Da.

Appendix 5: HBD:GAG binding studies

A5.1 Molecular dynamics simulations

In order to study the conformation adopted by the dimeric HBD2 plus Fx (D_{HBD2}+Fx) complex and to determine its theoretical CCS a theoretical study was undertaken, all calculations were performed by Dr Massimilano Porrini. The structure of Fx was taken from the PDB file of its complex with Antithrombin-S195A factor Xa (PDB 2GD4). Fx coordinates were taken from this PDB file and GLYCAM06 force field nomenclature was mapped onto them.[8] Sulfate groups were then attached at the correct N- and O- positions, topology paramaters and coordinate files were then created with the *tleap* module of Amber11.[9] The initial coordinates for HBD2 were taken from the crystal structure of the dimer (PDB 1FD3). Hydrogen atoms were added using *tleap* and the protonation states of the acidic and basic residues were assigned based on a solution pH of 7.

To create the complex the HBD2 dimer and Fx molecule were docked using Autodock 4.2 software[10] to predict the initial structure of the carbohydrate-protein complex. Autodock 4.2 features the implementation of flexible docking, allowing the rotation around diverse bonds. In the docking procedure used here all the bonds other than the glycosidic ones were allowed to rotate.

For determination of theoretical CCS monomer, dimer and complex structures were firstly minimised *in vacuo* with an "infinite" (999 Å) radial cut-off for the non-bonded interactions, and then gradually heated up to 300K, using a 1 fs time-step. The Amber99SB-ILDN[3] force field was used for the monomeric and dimeric species whilst the GLYCAM06[8] force field was used for pentasaccharide calculations. All MD simulations were run using the NAMD 2.9[11] simulation software. In all cases bonds involving hydrogen atoms were constrained at their equilibrium values using SHAKE algorithm.

For the dimeric simulations, in order to prevent the dissociation of the two monomers composing the dimer, a restraining harmonic potential was applied between the centres of mass of the monomers, with an equilibrium distance value equal to r_0=12.93 Å. This equilibrium distance is the average distance between the centres of mass of the two monomers determined from an 1 ns simulation of the dimer in water. A force constant of 10 kcal/mol/Å^2 was applied in order to keep the monomers together but to enable the dimer to explore the conformational space.

In vacuo production runs of 10 ns were carried out at constant temperature (300 K), with a 1 fs time-step. The rotationally averaged collision cross sections (CCS) were calculated with

the trajectory method of MOBCAL[1] one computation each 100 ps of the simulation time, for a total of 100 values.

A5.2 HBD2 plus Fx

Species	Expected mass (Da)	Observed mass (Da)
HBD2 Monomer (M_{HBD2})	4328.2	4327.4
HBD2 Dimer (D_{HBD2})	8656.5	8657.9
HBD2 Trimer (T_{HBD2})	12984.6	12987.1
HBD2 Dimer + Fx (D_{HBD2}+Fx)	10164.7	10165.9

Table A5.1: Expected and observed masses for all HBD2 and HBD2+Fx complexes, averaged over all observed charge states in the spectra in Chapter 6 Figure 6.2. The expected mass of fully desalted and protonated Fx is 1508.19 Da.

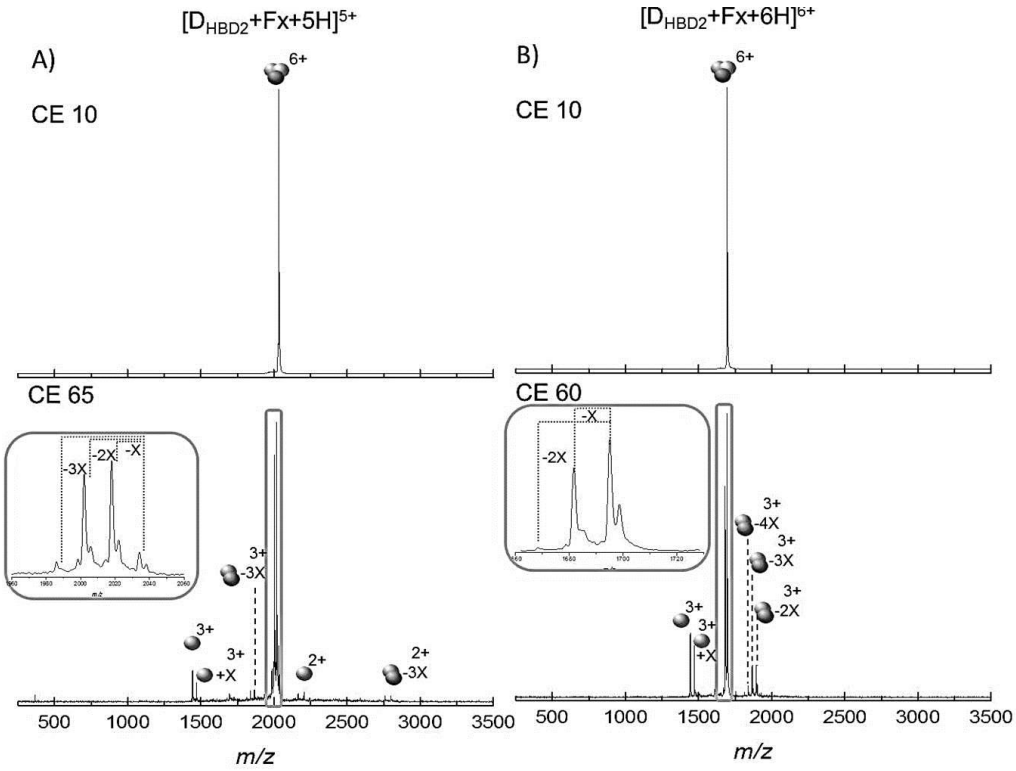

Figure A5.1: Spectra obtained for isolated species (top) and after CID (bottom) for A) $[D_{HBD2}+Fx+6H]^{6+}$ and B) $[D_{HBD2}+Fx+5H]^{5+}$. Acquired from a 1:1 (HBD2:Fx) solution prepared at 50 μM peptide concentration in 20 mM AmAc. Where X represents (SO_3).

Species	*Experimental CCS /Å²*	*Experimental CCS (absence of Fx) /Å²*
$[M_{HBD2}+3H]^{3+}$	582 ± 8	590
$[M_{HBD2}+4H]^{4+}$	589 ± 5	595
$[M_{HBD2}+5H]^{5+}$	641 ± 7	623
$[M_{HBD2}+6H]^{6+}$	683 ± 8	675
$[D_{HBD2}+6H]^{6+}$	928 ± 13	937

Table A5.2: Experimental CCS for all HBD2 monomeric and dimeric species, determined in the presence of Fx. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these. Experimental values obtained from the single repeat of HBD2 acquired in absence of Fx are also given here.

Species	Experimental CCS /Å²
$[D_{HBD2}+Fx+5H]^{5+}$	964 ± 6
$[D_{HBD2}+Fx+6H]^{6+}$	1023 ± 9
$[D_{HBD2}+Fx+7H]^{7+}$	1029 ± 3
$[D_{HBD2}+Fx+7H]^{7+}$	1070 ± 7
$[TE_{HBD2}+2Fx+10H]^{7+}$	1649 ± 17

Table A5.3: The experimental CCS for all HBD2+Fx complex species. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these.

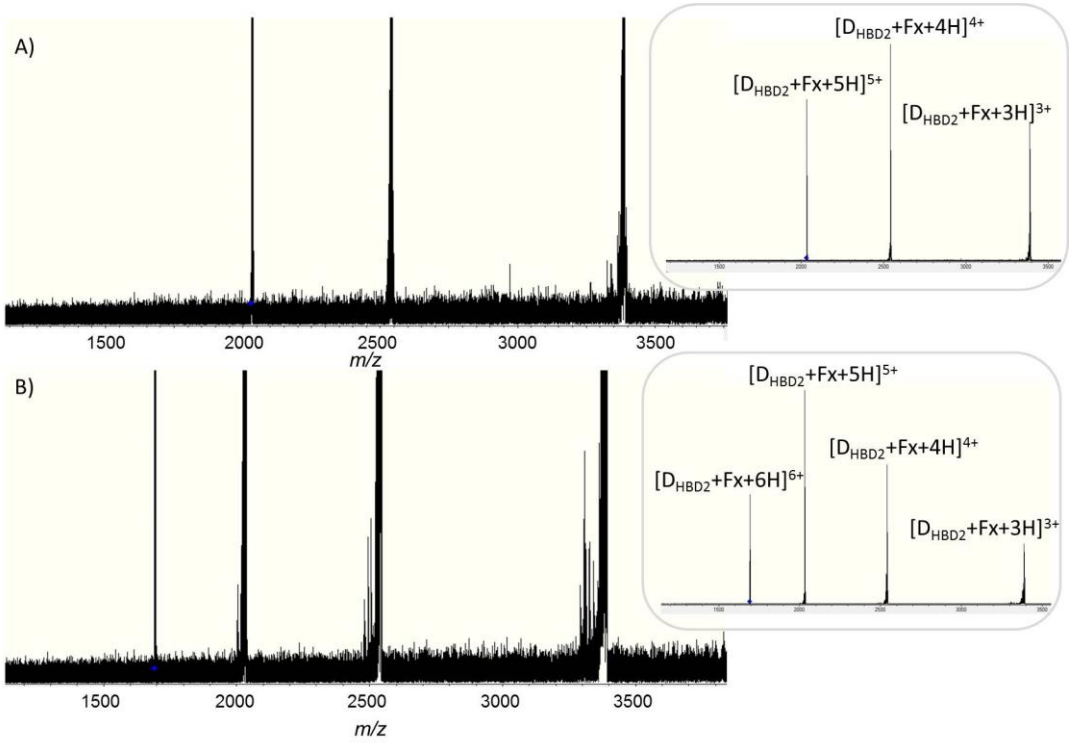

Figure A5.2: Mass spectra obtained following ECD on the 12T Apex Qe FTICR for HBD2 plus Fx expanded view to show ECD fragments, inserts show full range spectra. Acquired using nano-electrospray ionisation of a 1:1 sample solution of HBD2 plus Fx prepared at 30 μM peptide concentration in 20 mM AmAc. A) isolation of $[D_{HBD2}+Fx+5H]^{5+}$ B) isolation of $[D_{HBD2}+Fx+6H]^{6+}$.

Species	Experimental CCS /Å²
$[M_{HBD2}+4H]^{4+}$	593 ± 11
$[M_{HBD2}+5H]^{5+}$	661 ± 13
$[M_{HBD2}+6H]^{6+}$	698 ± 10
$[M_{HBD2}+7H]^{7+}$	747 ± 7
$[M_{HBD2}+8H]^{8+*}$	780
$[D_{HBD2}+Fx+6H]^{6+}$	1022 ± 7
$[D_{HBD2}+Fx+7H]^{7+}$	1064 ± 6
$[D_{HBD2}+Fx+8H]^{8+}$	1175 ± 24
$[D_{HBD2}+Fx+9H]^{9+}$	1256 ± 15

Table A5.4: The experimental CCS for all 'supercharged' HBD2 monomeric and dimeric species in addition to HBD2+Fx complex species. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these. * represents species observed in all repeats but only resolvable in one.

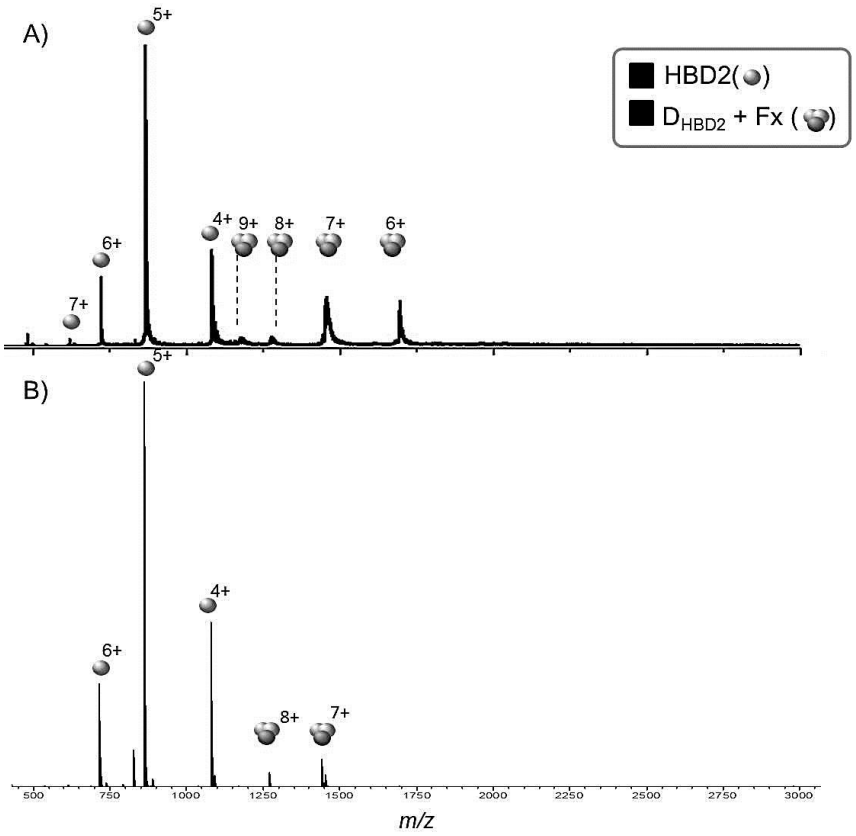

Figure A5.3: Spectra obtained for 1:1 sample solution of HBD2 plus Fx prepared in 20 mM AmAc + 1% *m*-NBA. A) 50 µM peptide concentration on the DT IM-MS instrument B) 30 µM peptide concentration on the 12T Apex Qe FTICR instrument.

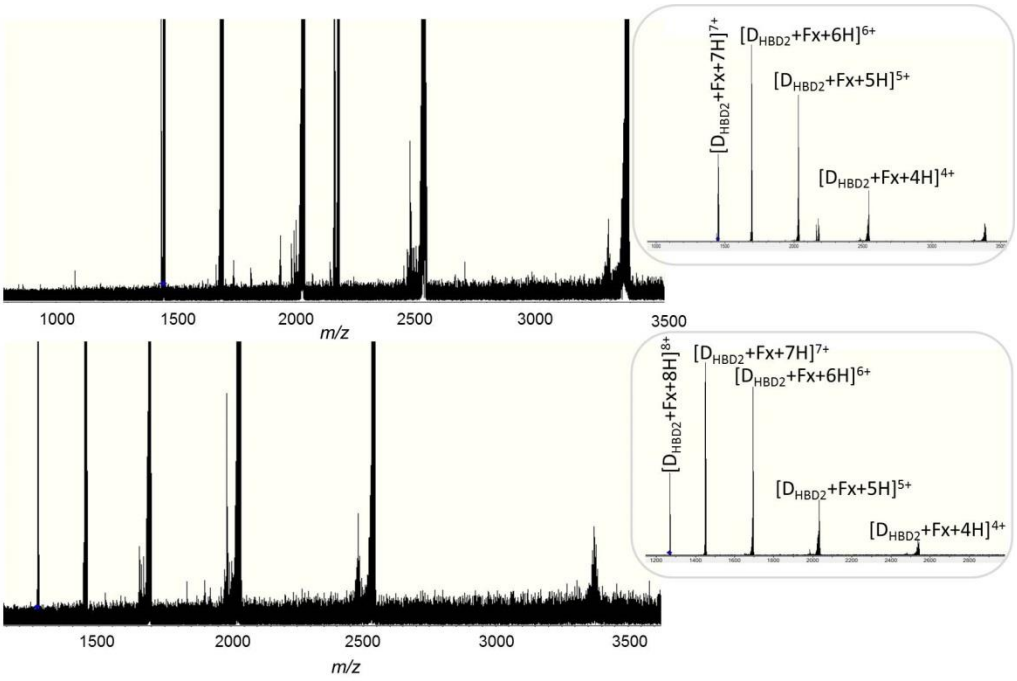

Figure A5.4: Mass spectra obtained following ECD on the 12T Apex Qe FTICR for HBD2 plus Fx zoomed in to show ECD fragments, inserts show full range spectra. Acquired using nano-electrospray ionisation of a 1:1 sample solution of HBD2 plus Fx prepared at 30 µM peptide concentration in 20 mM AmAc + 1% *m*-NBA. A) isolation of $[D_{HBD2}+Fx+7H]^{7+}$ B) isolation of $[D_{HBD2}+Fx+8H]^{8+}$.

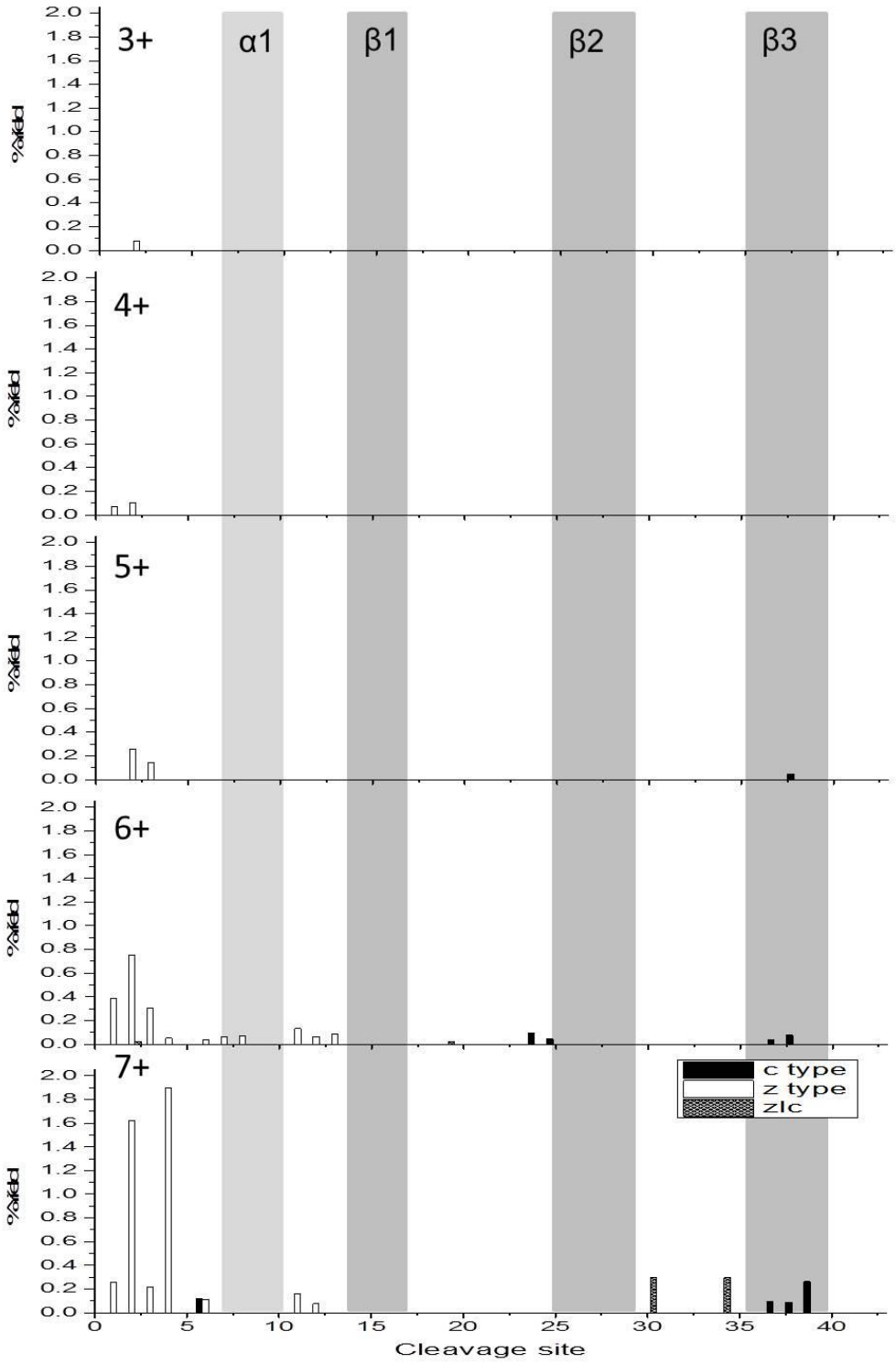

Figure A5.5: Percentage yields calculated for zic, c- and z-type ragments as a function of cleavage site for HBD2 over the charge states $[M_{HBD2}+3H]^{3+}$ to $[M_{HBD2}+7H]^{7+}$.

A5.3 HBD3 plus FX

Species	Expected mass (Da)	Observed mass (Da)
HBD3 Monomer (M_{HBD3})	5155.2	5155.9
HBD3 Dimer (D_{HBD3})	10310.4	10311.7
HBD3 Monomer + Fx (M_{HBD3}+Fx)	6663.4	6664.6
HBD3 Dimer + Fx (D_{HBD3}+Fx)	11818.6	11820.6
HBD3 Dimer +2 Fx (D_{HBD3}+2Fx)	13326.8	13327.7
HBD3 Trimer + Fx (T_{HBD3}+Fx)	16973.8	16975.3
HBD3 Trimer + 2Fx (T_{HBD3}+2Fx)	18482.0	18483.6
HBD3 Tetramer + 3Fx (T_{HBD3}+2Fx)	25145.3	25147.7
HBD3 Pentamer + 4Fx (P_{HBD3}+2Fx)	31808.7	31811.9

Table A5.5: Expected and observed masses for all HBD3 and HBD3+Fx complexes, averaged over all observed charge states in the spectra in Chapter 6 Figures 6.7 and 6.8. The expected mass of fully desalted and protonated Fx is 1508.19 Da.

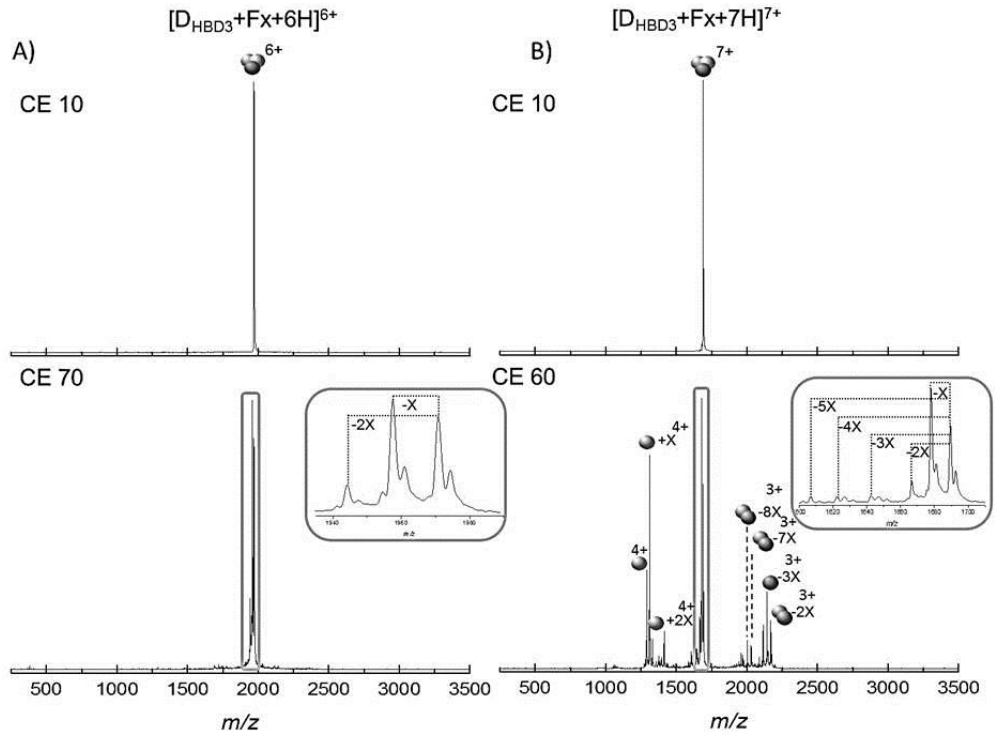

Figure A5.6: Spectra obtained for isolated species (top) and after CID (bottom) for A) $[D_{HBD3}+Fx+6H]^{6+}$ and B) $[D_{HBD3}+Fx+7H]^{7+}$. Acquired from a 2:1 (HBD3:Fx) solution prepared at 50 μM HBD3 concentration in 20 mM AmAc. Where X represents (SO_3).

Species	*Experimental CCS /Å²*	*Experimental CCS /Å² (absence of Fx)*
$[M_{HBD3}+4H]^{4+}$	746 ± 15	758
$[M_{HBD3}+5H]^{5+}$	760 ± 9	770
$[M_{HBD3}+6H]^{6+*}$	796	804
$[D_{HBD3}+8H]^{8+*}$	1164	n/a

Table A5.6: The experimental CCS for all HBD3 monomeric and dimeric species, determined in the presence of Fx. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these. Experimental CCS determined from a single repeat in the absence of Fx are also shown here.

Species	Experimental CCS /Å²	Theoretical CCS based on a spherical conformation/ Å²
$[M_{HBD3}+Fx+3H]^{3+}$	538 ± 20	571
$[M_{HBD3}+Fx+4H]^{4+}$	767 ± 27	
$[M_{HBD3}+Fx+5H]^{5+}$	854 ± 4	
$[D_{HBD3}+Fx+5H]^{5+}$	953 ± 21	806
$[D_{HBD3}+Fx+6H]^{6+}$	1174 ± 27	
$[D_{HBD3}+Fx+7H]^{7+}$	1265 ± 11	
$[D_{HBD3}+Fx+8H]^{8+}$	1330 ± 14	
$[D_{HBD3}+2Fx+6H]^{6+}$	966 ± 22	868
$[D_{HBD3}+2Fx+7H]^{7+}$	1187 ± 24	
$[D_{HBD3}+2Fx+8H]^{8+}$	1319 ± 89	
$[D_{HBD3}+2Fx+10H]^{10+}$	1470 ± 38	
$[T_{HBD3}+Fx+7H]^{7+}$	1415 ± 41	1006
$[T_{HBD3}+Fx+8H]^{8+}$	1544 ± 14	
$[T_{HBD3}+Fx+9H]^{9+}$	1707 ± 27	
$[T_{HBD3}+2Fx+8H]^{8+}$	1560 ± 66	1060
$[T_{HBD3}+2Fx+9H]^{9+}$	1659 ± 50	
$[T_{HBD3}+2Fx+10H]^{10+}$	1734 ± 28	
$[TE_{HBD3}+2Fx+10H]^{10+}$	1729 ± 17	1235
$[TE_{HBD3}+2Fx+12H]^{12+}$	1941 ± 27	
$[TE_{HBD3}+2Fx+14H]^{14+}$	2218 ± 70	
$[TE_{HBD3}+2Fx+16H]^{16+}$	2256 ± 93	
$[TE_{HBD3}+3Fx+11H]^{11+}$	1932 ± 44	1283
$[TE_{HBD3}+4Fx+14H]^{14+}$	2171 ± 38	1330
$[P_{HBD3}+3Fx+11H]^{11+*}$	2289	1441
$[P_{HBD3}+3Fx+12H]^{12+}$	2496 ± 8	
$[P_{HBD3}+4Fx+12H]^{12+}$	2468 ± 27	1485
$[P_{HBD3}+4Fx+13H]^{13+*}$	2472	
$[H_{HBD3}+2Fx+14H]^{14+}$	2484 ± 18	1546
$[H_{HBD3}+2Fx+16H]^{16+}$	2809 ± 93	
$[H_{HBD3}+2Fx+18H]^{18+}$	2886 ± 87	
$[H_{HBD3}+4Fx+16H]^{16+}$	2864 ± 42	1631
$[H_{HBD3}+4Fx+18H]^{18+}$	3030 ± 73	
$[H_{HBD3}+4Fx+20H]^{20+}$	3083 ± 45	
$[O_{HBD3}+6Fx+22H]^{22+}$	3355 ± 99	1978
$[DE_{HBD3}+6Fx+24H]^{24+*}$	4808	2225
$[DE_{HBD3}+6Fx+26H]^{26+*}$	4814	2295
$[DE_{HBD3}+8Fx+24H]^{24+}$	4465 ± 64	

Table A5.7: The experimental CCS for all HBD3+Fx complex species. * represents species observed in all repeats but only resolvable in one. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these. Theoretical CCS were determined as described in Chapter 6 section 6.3.3.3 based on a sphere of constant protein density. Where, monomeric (M), dimeric (D), trimeric (T), tetrameric (TE) pentameric (P), hexameric (H), octameric (O) and decameric (DE) HBD3 species are identified by letter codes.

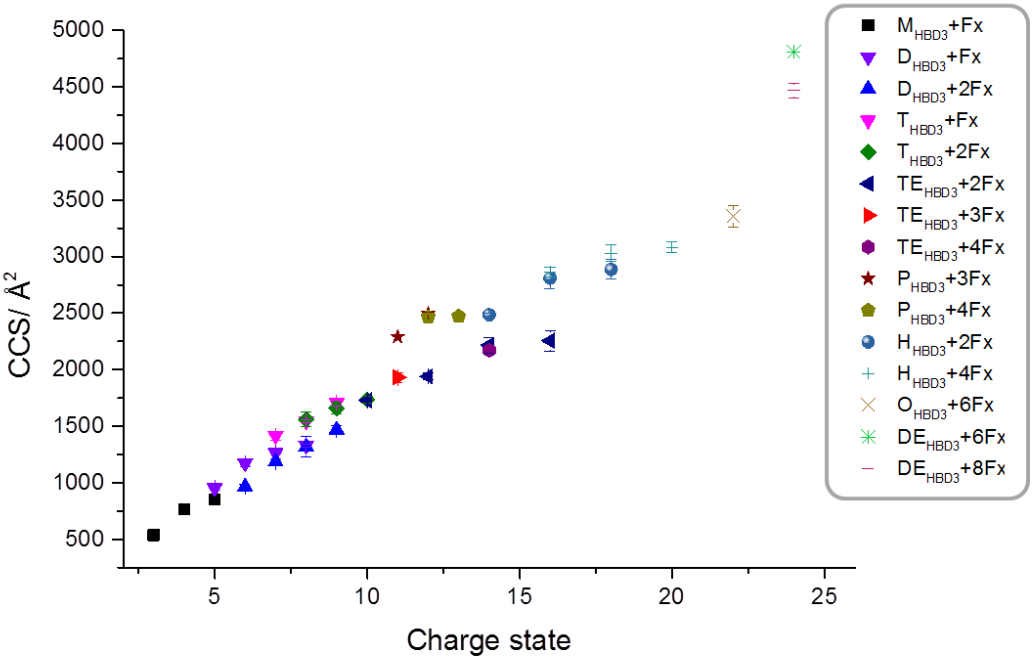

Figure A5.7: Average CCS determined for three repeats of all HBD3 plus Fx complexes, plotted as charge state versus CCS. Error bars represent the standard deviation between three repeats.

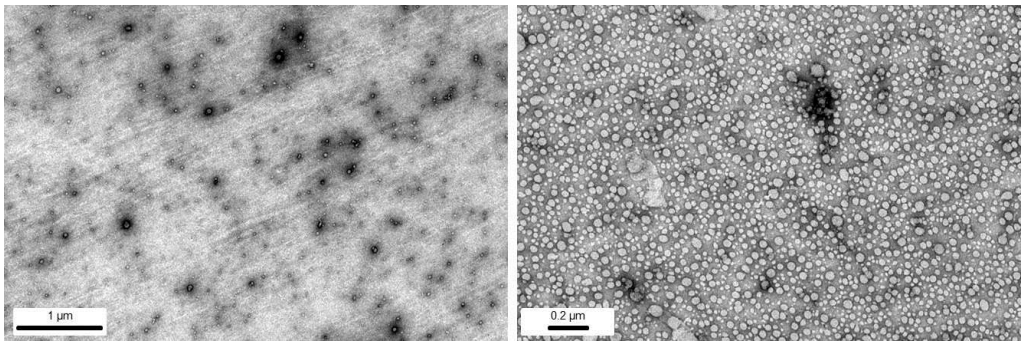

Figure A5.8: Representative TEM images for 65 µM HBD3 in 20 mM AmAc, stained with uranyl acetate

A5.4 HBD2 plus HBD3

Species	Expected mass (Da)	Observed mass (Da)
HBD2+HBD3 ($M_{HBD3+}M_{HBD3}$)	9483.4	9482.2
2HBD2+HBD3 ($D_{HBD3+}M_{HBD3}$)	13811.6	13810.3
3HBD2+HBD3 ($T_{HBD3+}M_{HBD3}$)	18139.8	18138.1
HBD2+2HBD3 ($M_{HBD3+}D_{HBD3}$)	14638.6	14637.5
HBD2+HBD3+Fx ($M_{HBD3+}M_{HBD3}$+Fx)	10991.59	10990.71
HBD2+2HBD3+Fx ($M_{HBD3+}D_{HBD3}$+Fx)	16146.79	16145.88

Table A5.8: Expected and observed masses for all HBD2+HBD3 and HBD2+HBD3+Fx complexes, averaged over all observed charge states in the spectra in Chapter 6 Figures 6.11 and 6.12. The expected mass of fully desalted and protonated Fx is 1508.19 Da.

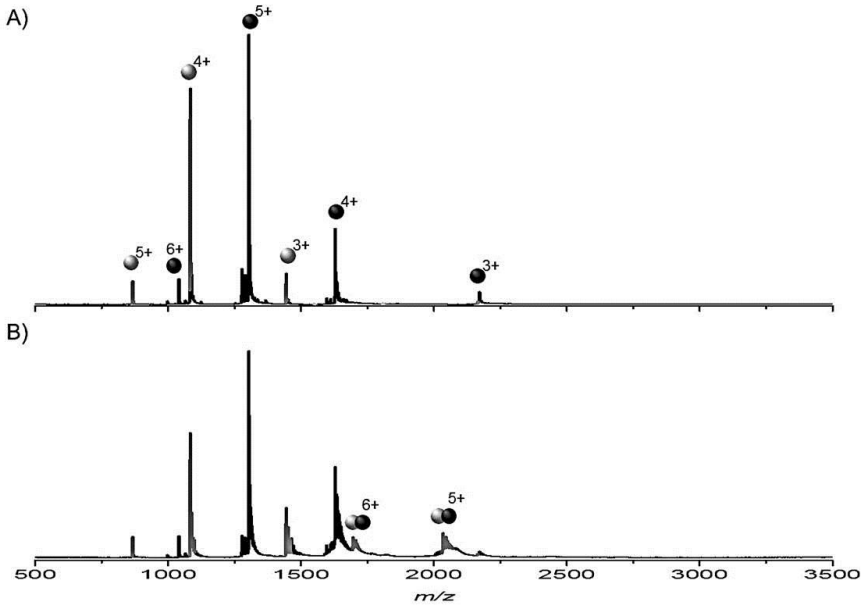

Figure A5.9: Mass spectrum of HBD2 plus BPTI A) prepared at a 1:1 ratio at 50 µM concentration B) HBD2 plus BPTI plus Fx prepared at 1:1:2 ratio at a 50 µM total protein concentration plus 50 µM Fx. In both cases samples were prepared in 20 mM AmAc.

Species	***Experimental CCS /Å²***	***Experimental CCS /Å² (determined in the presence of Fx)***
$[M_{HBD2}+M_{HBD3}+4H]^{4+*}$	1115	
$[M_{HBD2}+M_{HBD3}+5H]^{5+}$	1034 ± 18	1038 ± 23
$[M_{HBD2}+M_{HBD3}+6H]^{6+}$	1068 ± 5	1059 ± 22
$[M_{HBD2}+D_{HBD3}+7H]^{7+}$	1527 ± 4	
$[M_{HBD2}+D_{HBD3}+8H]^{8+}$	1639 ± 21	
$[D_{HBD2}+M_{HBD3}+6H]^{6+}$	1362 ± 24	
$[D_{HBD2}+M_{HBD3}+7H]^{7+}$	1252 ± 31	
$[T_{HBD2}+M_{HBD3}+8H]^{8+}$	1948 ± 24	
$[T_{HBD2}+M_{HBD3}+9H]^{9+*}$	1682	
$[T_{HBD2}+M_{HBD3}+9H]^{9+}$	1929 ± 16	
$[D_{HBD2}+D_{HBD3}+8H]^{8+*}$	1871	
$[TE_{HBD2}+D_{HBD3}+12H]^{12+}$	2421 ± 47	

Table A5.9: The experimental CCS for all HBD2 plus HBD3 complex species. * represents species observed in all repeats but only resolvable in one. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these. Peptide monomer (M), dimer (D), trimer (T) and tetramer (TE) are described by letter codes.

Species	Experimental CCS/ Å²	Experimental CCS/ Å² (determined from mixed peptide sample)
HBD2		
$[M_{HBD2}+3H]^{3+}$	582 ± 8	593±6
$[M_{HBD2}+4H]^{4+}$	589 ± 5	594±8
$[M_{HBD2}+5H]^{5+}$	641 ± 7	623±12
$[M_{HBD2}+6H]^{6+}$	683 ± 8	669±8
$[D_{HBD2}+6H]^{6+}$	928 ± 13	n/a
HBD3		
$[M_{HBD3}+4H]^{4+}$	746 ± 15	732±13
$[M_{HBD3}+5H]^{5+}$	760 ± 9	750±22
$[M_{HBD3}+6H]^{6+*}$	796	799±6
$[D_{HBD3}+8H]^{8+*}$	1164	n/a

Table A5.10: The experimental CCS for all HBD2 plus HBD3 complex species. * represents species observed in all repeats but only resolvable in one.

Species	Experimental CCS /Å²
$[M_{HBD2}+M_{HBD3}+Fx+5H]^{5+}$	1130 ± 29
$[M_{HBD2}+M_{HBD3}+Fx+6H]^{6+}$	1131 ± 9
$[M_{HBD2}+M_{HBD3}+Fx+7H]^{7+}$	1428 ± 43
$[M_{HBD2}+M_{HBD3}+Fx+7H]^{7+}$	1162 ± 12
$[M_{HBD2}+M_{HBD3}+Fx+8H]^{8+}$	1542
$[M_{HBD2}+M_{HBD3}+Fx+8H]^{8+*}$	1301 ± 5
$[M_{HBD2}+D_{HBD3}+Fx+9H]^{9+}$	1573 ±35
$[M_{HBD2}+D_{HBD3}+Fx+9H]^{9+}$	2163 ± 27
$[D_{HBD2}+D_{HBD3}+Fx+10H]^{10+}$	1859 ± 45

Table A5.11: The experimental CCS for all HBD2 plus HBD3 plus Fx complex species. The average values obtained from three different day repeats are reported here and errors are the standard deviations between these.

Appendix 6: Appendix References

(1) Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Chem. Phys.* **1998**, *108*, 2416-2423.
(2) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.
(3) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Struct. Funct. Bioinf.* **2010**, *78*, 1950-1958.
(4) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
(5) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(6) Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K. *University of California, San Francisco* **2012**.
(7) Roe, D. R.; Cheatham, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.
(8) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622-655.
(9) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.; Luo, R. C. W., W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg,; G. Seabra, I. K., K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell,; T. Steinbrecher, H. G., Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H.; Mathews, M. G. S., C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and; Kollman, P. A. *AMBER 11*; University of California: San Francisco, 2010
(10) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J. Comput. Chem.* **2009**, *30*, 2785-2791.
(11) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781-1802.